

Natural variation
at
FLOWERING LOCUS T
in
Arabidopsis thaliana

Inaugural-Dissertation
zur
Erlangung des Doktorgrades
der Mathematisch-Naturwissenschaftlichen Fakultät
der Universität zu Köln

Vorgelegt von
Emmanuel Tergemina
Aus Saint Martin d'Hères (Frankreich)

Köln, 2018



Max Planck Institute for
Plant Breeding Research

Die vorliegende Arbeit wurde am Max-Planck-Institut für Züchtungsforschung in Köln in der Abteilung für Entwicklungsbiologie der Pflanzen (Direktor Prof. Dr. G. Coupland) angefertigt.

Berichterstatter: Prof. Dr. George Coupland

Prof. Dr. Ute Höcker

Prüfungsvorsitzender: Prof. Dr. Maria Albani

Tag der Disputation: 10.07.2017



MAX-PLANCK-GESELLSCHAFT

Abstract

Flowering time is a complex trait that is crucial for fitness and regulated by seasonal changes and internal cues. In the annual plant *Arabidopsis thaliana* (Arabidopsis), two lifestyles have been described; winter accessions flower in response to long exposure to cold temperatures (vernalization) and single season accessions flower in a short period chiefly in response to long-days (LDs). The control of these two lifestyles converges at the level of transcriptional regulation of floral integrator genes, including the florigen encoding gene *FLOWERING LOCUS T* (*FT*). Indeed, *FT* expression, which is repressed prior vernalization in winter accessions, strongly accelerates flowering mainly in response to LDs but also to elevated temperatures.

The photoperiodic induction of *FT* expression requires a 5,7 kb region upstream of the start codon of *FT*. This area contains a proximal promoter and a distal enhancer phylogenetically conserved among the Brassicaceae. However, the distal enhancer is not required in the regulation of *FT* expression in the siliques, which has been proposed to influence related traits like floral reversion and seed dormancy.

While most of the flowering time quantitative trait loci (QTL) identified in Arabidopsis have been associated with variation at coding regions, QTLs closely linked to *FT* were exclusively mapped to non-coding regions. To test whether *FT* could be a target of selection to modulate flowering time, we dissected paired-end short read sequences from the 1001 genomes project for single nucleotide polymorphism and structural variation at the photoperiod control regions of *FT*. We identified rare *FT* haplotypes defined by natural variation in *cis*-regulatory elements of *FT* previously characterized in our group.

In the rosette leaves, prior the floral transition, allele-specific expression (ASE) of *FT* in F1 hybrids revealed *FT* haplotypes less responsive to both LDs and ambient temperatures. Interestingly, the ASE of *FT* in developing siliques of F1 hybrids showed a symmetric pattern, indicating that the causal natural variation does not affect the expression of *FT* in the siliques.

ASE in F1 hybrids and complementation experiments also revealed an *FT* haplotype, present in slow vernalization behavior accessions, relatively insensitive to inductive conditions in the rosette leaves. This last observation led us to investigate the effect of the florigen after extended vernalization. Surprisingly, the total number of leaves of *ft-10 tsf-1*

Abstract

double mutants is comparable to wild-type plants after prolonged periods of cold temperature.

These findings redefine the impact of *FT* on flowering time after extended periods of cold temperature. I propose that *FT* haplotypes less responsive to inductive conditions in the rosette leaves could accumulate in *Arabidopsis* populations with slow vernalization behaviors due to a relaxed purifying selection at *FT*.

Keywords:

Flowering time, *FLOWERING LOCUS T*, *cis*-regulatory evolution, natural variation, allele-specific expression (ASE).

Zusammenfassung

Die Blütezeit ist ein komplexes und für die biologische Fitness entscheidendes Charakteristikum einer Pflanze, das durch saisonale Veränderungen und innere Signale reguliert wird. Bei der einjährigen Pflanze *Arabidopsis thaliana* wurden zwei unterschiedliche Lebenszyklen beschrieben: winter accessions blühen erst nachdem sie längere Zeit kalten Temperaturen (Vernalisation) ausgesetzt waren und single season accessions blühen nach kurzer Zeit unter Langtag Bedingungen (long-day, LD).

Die Kontrolle dieser zwei Lebenszyklen steht in Zusammenhang mit der transkriptionellen Regulierung von *FLOWERING LOCUS T (FT)*, einem Florigen entkodierenden Gen. Tatsächlich beschleunigt die Expression von *FT*, die in winter accessions vor der Vernalisation gehemmt wird, das Blühen überwiegend im LD, aber auch unter erhöhten Temperaturen im short-day. Die fotoperiodische Aktivierung von *FT* beruht auf einer 5,7 kb langen Region oberhalb des Start-Codons von *FT*. Dieser Bereich enthält einen proximalen Promoter und einen distalen Enhancer, die innerhalb der Brassicaceae konserviert sind mit einer offenen Chromatinstruktur. Der distale Enhancer ist für die Regulation der *FT* Expression in den Schoten jedoch nicht erforderlich. Bisher wurde angenommen, dass dadurch die verwandten Merkmale der floral reversion und die Samenruhe beeinflusst werden. Während die meisten quantitative trait loci (QTL) für die Blütezeit, die in *Arabidopsis* identifiziert wurden, mit Variationen in codierenden Regionen in Verbindung gebracht wurden, stehen QTL um *FT* ausschließlich mit der Promoterregion in Verbindung. Um zu testen, ob *FT* ein Selektionsziel darstellt um die Blütezeit zu modulieren, haben wir Illumina paired-end Sequenzen des 1001 Genoms-Projekts nach Punktmutationen und struktureller Variation in den Bereichen, die für fotoperiodische Regulation verantwortlich sind, durchsucht.

Hierbei haben wir seltene *FT* Haplotypen entdeckt, die durch natürliche Variation in bereits von unserer Gruppe beschriebenen *cis*-regulatorischen Elementen des *FT* Gens definiert werden. Analyse der Allelspezifischen Expression (ASE) von *FT* in Rosettenblättern von vegetativen F1-Hybriden führte zur Entdeckung von *FT* Haplotypen, die weniger auf LD und Umgebungstemperatur reagierten. Interessanterweise zeigte ASE von *FT* ein symmetrisches Muster in heranwachsenden Schoten von F1 Hybriden, was darauf hindeutet, dass die ursächliche natürliche Variation die Expression von *FT* in den Schoten nicht beeinflusst. ASE in F1-Hybriden identifizierten einen weiteren *FT* Haplotypen, der

Zusammenfassung

in „slow vernalizing accessions“ existiert und in den Rosettenblättern verhältnismäßig schwach auf induktive Bedingungen reagiert.

Diese letzte Beobachtung führte uns dazu den Effekt von *FT* nach umfangreicher Vernalisation zu untersuchen. Erstaunlicherweise ist die Gesamtblattzahl der *ft-10 tsf-1* Doppelmutante vergleichbar mit Pflanzen des Wildtyps nach längerer Zeit bei niedriger Temperatur.

Diese Ergebnisse verändern unser Verständnis der Auswirkung von *FT* auf die Blütezeit nach längerer Zeit bei niedriger Temperatur. Meiner Meinung nach kann eine Anhäufung von wenig auf induktive Bedingungen reagierenden *FT* Haplotypen in *Arabidopsis* Populationen mit langsamem Vernalisationsverhalten auf verminderte negative Selektion hindeuten.

Stichworte:

Blütezeit, *FLOWERING LOCUS T*, *cis*-regulatory evolution, Allelspezifische Expression (ASE).

Table of content

Abstract	I
Zusammenfassung.....	III
Table of content.....	V
1 Introduction.....	1
1.1 The regulation of flowering time: how to flower at the right time?.....	1
1.1.1 Flowering in response to external cues	1
1.1.1.1 Photoperiod.....	1
1.1.1.2 Temperature.....	3
1.1.1.2.1 Vernalization.....	3
1.1.1.2.2 Ambient temperature.....	4
1.1.2 Flowering in response to internal cues	5
1.1.2.1 Aging.....	5
1.1.2.2 Gibberellic acid levels.....	6
1.1.2.3 Autonomous pathway.....	6
1.2 Integration of flowering signals at the level of <i>FT</i> transcriptional control.....	7
1.2.1 The <i>FT</i> locus	7
1.2.2 Induction of <i>FT</i> expression.....	8
1.2.2.1 <i>Cis</i> -regulatory elements at <i>FT</i> involved in the photoperiod-mediated response.....	8
1.2.2.1.1 <i>Cis</i> -regulatory regions at <i>FT</i> involved in the photoperiod-mediated response.....	8
1.2.2.1.2 Transcription factors binding sites at <i>FT</i> involved in the photoperiod-mediated response...	9
1.2.2.2 <i>Cis</i> -regulatory elements at <i>FT</i> involved in the ambient temperature-mediated response	12
1.2.3 Repression of <i>FT</i> expression	13
1.3 Natural variation of flowering time in Arabidopsis	13
1.3.1 Arabidopsis as a model to study natural variation	13
1.3.3 Tools to study natural variation in Arabidopsis.....	16
1.3.3.1 Identification of the gene.....	16
1.3.3.2 Identification of the causal natural variation	18
1.3.4 Natural genetic variation affecting flowering time in Arabidopsis	20
1.3.5 Difference of flowering time due to natural variation at <i>FT</i>	21
2. Aim of the study.....	21
3. Results.....	27
3.1 Natural variation in the coding sequence of <i>FT</i>	27
3.2 Natural variation in the non-coding sequence of <i>FT</i>	29
3.2.1 Natural variation in the promoter length of <i>FT</i>	29
3.2.2 Natural variation in the photoperiod control regions of <i>FT</i>	32

Table of content

3.2.3 Transposable element polymorphisms at <i>FT</i>	41
3.3 Characterization of the candidate non-coding variant <i>FT</i> haplotypes	44
3.3.1 Allele-specific expression of <i>FT</i> in F1 hybrids.....	44
3.3.1.1 Allele-specific expression of <i>FT</i> in the rosette leaves.....	46
3.3.1.1.1 LD photoperiod.....	46
3.3.1.1.2 Ambient temperature.....	47
3.3.1.2 Allele-specific expression of <i>FT</i> in the siliques.....	49
3.3.2 Complementation assay.....	50
3.4 Flowering behavior of the parental accessions	54
3.5 Characterization of the impact of <i>FT</i> after extensive cold exposure	56
3.6 Characterization of winter <i>Arabidopsis</i> accessions.....	57
4. Discussion	59
4.1 Natural variation at <i>FT</i> in <i>Arabidopsis</i>	59
4.2 Functional characterization of <i>FT</i> haplotypes with natural variation in the photoperiod control regions of <i>FT</i> in the rosette leaves.....	60
4.2.1 Characterization in F1 hybrids.....	60
4.2.2 Complementation experiments	62
4.3 The photoperiod and thermosensory induction of <i>FT</i> expression share CREs.....	63
4.4 <i>FT</i> haplotypes less sensitive to inductive conditions are affected in a tissue specific manner	63
4.5 Footprint of relaxed selection of <i>FT</i> in accessions with a slow vernalization behavior?	64
4.6 Florigen has little impact on the floral transition after extensive periods of cold temperature	65
4.7 The detection of functionally differentiated <i>FT</i> haplotypes is a complex scenario	66
5. Material and methods	68
5.1 Screening for natural variation at <i>FT</i>.....	68
5.1.1 Single nucleotide polymorphisms	68
5.1.2 Structural variations	68
5.1.2.1 Mapping	68
5.1.2.2 Structural variations detection	69
General	69
Functional characterization.....	69
5.1.3 Phylogenetic footprinting	70
5.1.4 Screening for candidate accessions	70
5.2 Plant Growth	70
5.3 Allele-specific expression analysis.....	71
5.4 Generation of transgenic lines.....	72

5.4.1 PCR	72
5.4.2 Transformation of <i>Escherichia coli</i>	72
5.4.3 Plasmid construction	72
5.4.3.1 Generation of Gateway® ENTRY vectors	72
5.4.3.2 Generation of Gateway® destination vectors with a selectable marker.....	73
5.4.3.3 Generation of binary vectors	74
5.4.4 Plant transformation and transgenic lines.....	74
5.4.5 Screening for tandem repeat insertions	75
5.5 Flowering time measurement.....	75
5.6 Statistical analysis.....	76
5.7 Geographical distribution	76
References.....	77
Appendix.....	87
Abbreviations	135
General.....	135
Genes and proteins	138
Amino acids	140
Acknowledgements.....	141
Erklärung	142
Lebenslauf	143

1 Introduction

1.1 The regulation of flowering time: how to flower at the right time?

Flowering time corresponds to the transition from vegetative growth to reproductive development and is a crucial prerequisite for successful sexual reproduction. The correct timing of the floral transition allows seed development in optimal conditions and therefore contributes to fitness. This developmental change is modulated by different exogenous signals including biotic and abiotic factors, and endogenous signals comprising developmental age and amount of gibberellic acid (GA). Isolation of flowering time mutants in the model plant species *Arabidopsis thaliana*, called hereafter *Arabidopsis*, significantly contributed to the understanding of flowering time and defined the genetic pathways that regulate the floral transition in response to these different signals.

1.1.1 Flowering in response to external cues

1.1.1.1 Photoperiod

Flowering plants developed the ability to sense seasonal changes in day length, a process called photoperiodism, to adapt their flowering performance to their native environment. According to their behavior in response to length of the day, plants can be classified into three groups. Long-day plants flower faster when exposed to more than a critical day length (late spring, early summer), short-day plants flower faster when exposed to less than a critical day length (fall, winter) and neutral-day plants are not influenced by the length of the day (Garner and Allard 1920).

Photoperiodism involves critical day length perception in the leaves by essential circadian clock-mediated components that accelerate the floral transition in the shoot apical meristem (SAM) through multiple transcriptional regulatory cascades, post-translational modifications and protein movement. The model plant *Arabidopsis*, which has been used to understand the foundation of flowering time, is a facultative long-day plant distributed among the northern hemisphere.

Introduction

The key floral promoting genes involved in photoperiod flowering include *FLAVIN KELCH FACTOR 1 (FKF1)*, *GIGANTEA (GI)*, *CONSTANS (CO)* and *FLOWERING LOCUS T (FT)*. FKF1 and GI proteins are two circadian clock components able to interact exclusively in LDs to ultimately activate *CO* transcription (Sawa et al. 2007). The post-translational regulation of *CO* is also primordial in this photoperiodic process. In the dark, the SUPPRESSOR OF PHYTOCHROME (SPA1) protein acts together with the CONSTITUTIVE PHYTOMORPHOGENIC 1 (COP1) ubiquitin-ligase to mediate the degradation of CO protein by the 26S proteasome (Valverde et al. 2004; Jang et al. 2008). However, in the morning, Phytochrome B (PhyB), a photoreceptor sensitive to red light, leads the repression of CO protein abundance through a COP1-independent mechanism (Valverde et al. 2004; Jang et al. 2008). During the day, CRYPTOCHROME 2 (CRY2) and PHYTOCHROME A (PHYA), two photoreceptors respectively sensitive to blue and far-red lights, prevent CO degradation by repressing the SPA1-COP1 complex (Yang et al. 2000; Wang et al. 2001; Zuo et al. 2011). Thus, the combination of transcriptional and post-translational regulation of *CO* leads to a peak of CO protein at the end of the day. CO protein contains two zinc-binding B-boxes domains and a CO CONSTANS-Like TIME OF CAB1 (CCT) domain (Putterill et al. 1995). The CCT domain shares similarities with NUCLEAR FACTOR-Y (NF-Y) proteins (Siefers et al. 2009). CO and the NF-Ys have been shown to act in concert as a trimeric complex to induce the expression of *FT*, and its paralog *TWIN SISTER OF FT (TSF)*, in the phloem companion cells (Samach et al. 2000; Yamaguchi et al. 2005; Wenkel et al. 2006). FT is a phosphatidylethanolamine-binding protein (PEBP), which shares homology to Raf Kinase Inhibitor Proteins (RKIPs), and has been shown to be a key component of the florigen signal transmitted from the leaves through the phloem to the SAM (Kotake et al. 2003; Corbesier et al. 2007). This transport requires FT-INTERACTING PROTEIN 1 (FTIP1), an endoplasmic reticulum (RE) membrane protein (Liu et al. 2012). Once in the SAM, FT interacts with FLOWERING LOCUS D (FD), a basic leucine zipper domain transcription factor, to induce the expression of floral meristem identity genes like *LEAFY (LFY)*, *APETALA 1 (API)* or *SUPPRESSOR OF CONSTANS 1 (SOC1)* and ultimately accelerate the floral transition (Schmid et al. 2003; Abe et al. 2005; Yoo et al. 2005; Searle et al. 2006).

1.1.1.2 Temperature

1.1.1.2.1 Vernalization

Many plants require prolonged exposure to cold temperature to accelerate the floral transition, a process called vernalization. In *Arabidopsis*, two lifestyles have been described; single season accessions that flower faster mainly in response to LDs, such as the reference accession Col-0, and winter accessions that flower faster in response to vernalization. Genetic analysis of *Arabidopsis* accessions that differ in their vernalization response identified the core genetic components behind these two different lifestyles. Winter accessions carry often active alleles of the two floral repressors *FRIGIDA* (*FRI*) and *FLOWERING LOCUS C* (*FLC*), whereas single season accessions contain natural hypomorphic or even amorphic allele of at least one of these two genes (Johanson et al. 2000; Shindo et al. 2005; Werner et al. 2005). Before winter, the coiled-coil protein *FRI* promotes *FLC* expression (Johanson et al. 2000). *FLC* protein together with SHORT VEGETATIVE PHASE (*SVP*) belong to the MADS-domain transcription factor family and have been proposed to act as a heterotetrameric complex to block photoperiod flowering by the repression of floral integrator genes expressed in the leaves and the SAM like *FT* and *SUPPRESSOR OF OVEREXPRESSION OF CONSTANS 1* (*SOC1*), respectively (Michaels and Amasino 1999; Searle et al. 2006; Li et al. 2008).

The repression of *FLC* expression during vernalization is highly complex and requires several independent regulatory steps. After only two to three weeks of cold treatment, the downregulation of *FLC* expression reaches saturation in concert with an upregulation of the long noncoding RNA (lncRNA) *COOLAIR* which is a noncoding antisense transcript of *FLC* (Swiezewski et al. 2009). Afterwards, extended exposure to cold temperature promotes epigenetic silencing of *FLC*. Indeed, the cold-induced repression of *FLC* expression is maintained throughout the subsequent mitotic cell divisions until resetting during embryogenesis (Sheldon et al. 2008; Crevillen et al. 2014). Quite recently, the first intron of *FLC* has been shown to contain *cis*-regulatory elements (CREs) required for the epigenetic silencing of *FLC* through the recruitment of the Polycomb machinery (Qüesta et al. 2016; Yuan et al. 2016). This vernalization-mediated repression of *FLC* expression has been shown to correlate with accumulation of the tri-methylation at lysine 27 of histone

H3 (H3K27me3), a repressive histone methylation mark characteristic of Polycomb group (PcG) silencing and deposited by the PcG repressive complex 2 (PRC2) (Angel et al. 2011). This prolonged cold treatment induces the formation of plant homeodomain (PHD)-PRC2 complexes composed of the core component of PRC2 and the plant homeodomain proteins including VERNALIZATION 5 (VRN5) and VERNALIZATION INSENSITIVE 3 (VIN3) (De Lucia et al. 2008). Additionally, several lncRNAs, comprising COLDAIR and COLDWRAP, have also been suggested to be involved in the recruitment of PRC2 to *FLC* (Heo and Sung 2011; Kim and Sung 2017).

1.1.1.2.2 Ambient temperature

In addition to cold temperature sensitivity, plants developed the ability to accelerate the floral transition in response to ambient temperatures. This thermosensory response is able to overcome non-inductive photoperiod conditions. The main genes involved in this floral promotion include *SVP*, *FLOWERING LOCUS M (FLM)* and *PHYTOCHROME-INTERACTING FACTOR 4 (PIF4)*.

The MADS-domain transcription gene *FLM* plays a central role in this process. *FLM* generates at least four spliced variants including *FLM-β* and *FLM-δ*, the most predominant isoforms in the reference accession Col-0 (Scortecci et al. 2001). These two isoforms have been shown to accumulate at different temperatures; the abundance of the *FLM-β* isoform decreases from 16°C to 27°C while the abundance of the *FLM-δ* isoform increases. At low temperature, FLM-β protein interacts with SVP to repress floral integrator genes, such as *FT* and *SOC1*, to consequently repress the floral transition. However, in warmer ambient temperature, SVP protein has been shown to be rapidly degraded by the 26S proteasome (Lee et al. 2013). The combination of warm ambient temperature mediated degradation of SVP protein together with a reduction of *FLM-β* transcript leads ultimately to a decrease in the repressive SVP-FLM-β complex abundance and therefore relieve the repression of *FT* and *SOC1* expression. Additionally, FLM-δ protein has also been suggested to sequester SVP in an inactive complex during this thermosensory response (Pose et al. 2013).

In parallel to FLM, the basic helix-loop-helix (bHLH) transcription factor PIF4 has also been proposed to mediate this thermosensory response (Kumar et al. 2012). Despite the

absence of a distinct flowering phenotype in *pif4* null allele mutants exposed to LDs, the impact of *PIF4* appears in ambient temperature under short days (SDs) (Koini et al. 2009; Fernández et al. 2016). The *ft-10* null allele strongly delays the floral transition in these conditions and indicates a thermosensory response strongly dependent on *FT* (Balasubramanian et al. 2006). *PIF4*, in addition to CO, has been shown to directly promote the expression of *FT* and *TSF* in these conditions. However, despite a comparable flowering time compared to LDs, the level of *FT* messenger RNA (mRNA) in wild-type plants is radically lower and brings into question the SAM sensitivity to the florigen (Fernández et al. 2016).

1.1.2 Flowering in response to internal cues

1.1.2.1 Aging

Plants acquire the competence to flower in response to different environmental signals after the vegetative transition from juvenile to adult phase marked by the formation of abaxial trichomes on the rosette leaves (Chien and Sussex 1996; Telfer et al. 1997). This age-related competence to flower was first described in perennial species since *Arabidopsis* responds to inductive conditions particularly early after germination (Bergonzi and Albani 2011; Huijser and Schmid 2011). The microRNA156 (miR156) and microRNA172 (miR172) have been shown to play a significant role in this developmental transition. The abundance of miR156 and miR172 has opposite temporal patterns and effects on vegetative phase change. The miR156 has been shown to be highly abundant at the juvenile phase and to repress the mRNA of several *SQUAMOSA PROMOTER BINDING PROTEIN LIKE* (*SPL*) genes (Wang et al. 2009). *SPL* transcription factors, *SPL9* and *SPL15* in particular, promote directly *miRNA172* expression. The miR172 has been demonstrated to target the mRNA of the *APETALA2* (*AP2*)-like genes including the floral repressors *SCHLAFMUTZE* (*SMZ*), *SCHNARCHZAPFEN* (*SNZ*), *TARGET OF EAT 1* (*TOE1*) and *TOE2* (Aukerman and Sakai 2003; Mathieu et al. 2009). Increase in miR172 abundance results in the downregulation of *AP2*-like genes expression which subsequently promotes the switch to adult vegetative phase and indirectly promotes the expression of floral integrator genes like *FT*. Indeed, *FT* expression has been proposed to be repressed by *SMZ* (Mathieu et al. 2009). In parallel, *SPL* proteins were shown to induce the expression of

Introduction

floral identity genes like *LFY*, *API* and *FRUITFULL (FUL)* (Yamaguchi et al. 2009). Additionally, *SPL15* has been shown to form a transcriptional regulatory complexed with *SOC1* to promote *FUL* expression (Hyun et al. 2016). This finding indicates that plants could acquire the competence to flower without floral inductive environmental conditions if they pass the juvenile phase and have high levels of GA.

1.1.2.2 Gibberellic acid levels

Gibberellic acid (GA) is a crucial plant growth regulator involved in different developmental processes. Analysis of mutants impaired in GA biosynthesis revealed a substantial impact of GA in non-inductive photoperiods. Indeed, the *gal-3* mutant exposed to LDs shows a relatively weak late-flowering phenotype compared to Col-0 but fails to flower in SDs, suggesting that photoperiod flowering can overcome the loss of GA signaling (Wilson et al. 1992). However, GA signaling has been shown to be closely linked to photoperiodic flowering since it promotes the expression of *FT* and *TSF* (Porri et al. 2012). Additionally, GA also influences the competence to flower through the promotion of *SPL* genes expression (Hyun et al. 2016).

1.1.2.3 Autonomous pathway

Mutants screen also revealed late flowering mutants not influenced by the photoperiod and barely affected by the vernalization (Koornneef et al. 1991). The genes identified were classified in the autonomous pathway and include *FLOWERING LOCUS CA (FCA)*, *FLOWERING LOCUS Y (FY)*, *FLOWERING LOCUS PA (FPA)* and *FLOWERING LOCUS KH DOMAIN (FLK)* involve in RNA regulatory processes, and *FLOWERING LOCUS D (FLD)*, *LUMINIDEPENDENS (LD)* and *FLOWERING LOCUS VE (FVE)* involved in chromatin modification. All of these components share in common *FLC* as a target (Marquardt et al. 2006).

1.2 Integration of flowering signals at the level of *FT* transcriptional control

The different flowering time pathways present several cross-talks and converge at the level of transcriptional regulation of floral integrator genes including *FT* (Pin and Nilsson 2012). *FT* has been shown to be primordial in inductive conditions including photoperiod and ambient temperature. A loss of function of *FT* leads to a late flowering phenotype in LDs/21°C and SDs/27°C whereas its overexpression induces an early flowering phenotype independent of external and internal signals (Koornneef et al. 1991; Kardailsky et al. 1999; Kobayashi et al. 1999; Fernández et al. 2016). The closest homolog of *FT*, *TSF*, also produces a functional protein. However, a null allele of *TSF* has a marginal impact on the floral transition compared to wild-type plants which brings up FT at the head of the florigen signal at least in the Col-0 background. *TSF* expression has been reported to be considerably lower than *FT* during photoperiodic flowering. This weaker *TSF* expression has been suggested to be due to the presence of a transposable element (TE) located downstream of the 3' untranslated region (UTR) of *TSF*, which influences the chromatin environment of *TSF* itself (Yamaguchi et al. 2005).

1.2.1 The *FT* locus

FT is conserved in flowering plants (Ho and Weigel 2014). Considering the small intergenic regions in Arabidopsis, *FT* is rather isolated from the neighboring genes and flanked by large intergenic sequences (~7,5 kb upstream and downstream of *FT*). The identification of active *cis*-regulatory elements (CREs) is essential for the understanding of the transcriptional regulation of a gene in a spatial and temporal-specific manner. The classical approach to detect the active CREs consists in the functional characterization of phylogenetically conserved sequences. Our group revealed by phylogenetic shadowing three highly conserved regions upstream of *FT* named *Block A*, *Block B*, and *Block C* (Adrian et al. 2010), and one downstream of *FT* called *Block E* (Johan Zicola, Thesis). Interestingly, *Block A* is also conserved at the *TSF* locus (Adrian et al. 2010). Chromatin immunoprecipitation (ChIP) revealed that the repressive mark H3K27me3 is widely spread throughout the *FT* locus with the exception of *Block C* and *Block E* (Adrian et al. 2010). Additionally, genome-wide high-resolution maps of DNase I hypersensitivity (DH) reveal that these phylogenetically conserved CREs correlate with open chromatin regions and are more likely accessible to transcription factors (Zhang et al. 2012). Transcription factor

binding sites (TFBSs) represent the functional unit of CREs and are therefore the key to the understanding of spatiotemporal gene regulation. The canonical approach relies on the combination of methods like DNA footprinting and ChIP, done respectively *in vitro* and *in vivo*, to test the affinity of a transcription factor to short sequences phylogenetically conserved. Ultimately, the proof for the identification of active TFBSs *in planta* should rely on complementation tests.

1.2.2 Induction of *FT* expression

1.2.2.1 *Cis*-regulatory elements at *FT* involved in the photoperiod-mediated response

1.2.2.1.1 *Cis*-regulatory regions at *FT* involved in the photoperiod-mediated response

To identify the CREs located upstream of *FT* required for its photoperiodic induction expression in the rosette leaves, different constructs consisting of various sequence lengths upstream of the translation start site (TSS) of *FT* fused to the *FT* cDNA were generated to test their ability to complement the late flowering phenotype of the *ft-10* mutant. The 5.7 kb *FT* promoter construct (*5.7kbFTp::FTcDNA*) rescues the *ft-10* mutant phenotype in LDs while the 5.2 kb, the 4.0 kb, and the 1.0 kb *FT* promoters fail (Adrian et al. 2010; Liu et al. 2014a). This trend is also true in histochemical β -Glucuronidase (GUS) localization assays; only the 5.7 kb *FT* promoter can drive *GUS* reporter gene expression in the vasculature of the first true leaves of wild-type plants exposed to LDs in a comparable manner to spatial *FT* expression pattern already described (Takada and Goto 2003; Adrian et al. 2010). Furthermore, our group reported that methylation at *Block C*, induced by the expression of inverted repeats (IR), reduced the expression of *FT* leading consequently to a late flowering phenotype compared to wild type plants (Johan Zicola, Thesis). Moreover, overexpression of *CO* under the control of the *35S Cauliflower Mosaic Virus (CaMv)* promoter increases the expression of the *GUS* reporter gene driven by a 5.7 kb *FT* promoter in photoperiod independent manner (Adrian et al. 2010).

Taken together, these results demonstrated that *Block C*, located between 5.7 kb and 5.2 kb, is essential for the CO-mediated photoperiodic induction of *FT* expression in the rosette leaves.

Additionally, a construct consisting of *Blocks A* and *C* fused to the coding sequence of *FT* (*C+Ap::FTcDNA*) rescues the *ft-10* mutant phenotype while a construct made of *Block C* fused to the minimal promoter of the Nopaline Synthase (NOS) gene (*BlockC+minNOSp::FTcDNA*) fails (Liu et al. 2014a). Consistent with these results, *C+Ap* has the ability to drive *GUS* expression in a comparable manner to spatial *FT* expression pattern already published, while the *BlockC+minNOSp* does not (Liu et al. 2014a). Taken together, the data indicate that the CO-mediated photoperiodic induction of *FT* expression in the leaves requires *Block A* as a proximal promoter and *Block C* as a distal enhancer. Interestingly, *Block B*, located between *Block A* and *Block C*, seems to be not essential at least in this developmental process (Liu et al. 2014a).

Recently, our group showed that transgenic plants, expressing IR targeted to *Block E*, showed a reduced level of *FT* mRNA and a late flowering phenotype compared to wild plants in photoperiod inductive conditions. Taken together, the data defined *Block E* as a novel enhancer involved in the photoperiodic induction of *FT* expression.

1.2.2.1.2 Transcription factors binding sites at *FT* involved in the photoperiod-mediated response

Phylogenetic footprinting at *Blocks A* and *C* revealed several short sequences highly conserved among the Brassicaceae likely to be TFBSs under selective constraints. At *Block A*, our group previously identified four short conserved “shadow” sequences annotated as *Shadow 1* to *4* (*S1* to *S4*) and two palindromic sequences flanking *S3* called *P1* and *P2*. These sequences were first tested in a transient expression assay using constructs made of the Luciferase (Luc) reporter gene under the control of a 1 kb *FT* promoter with mutations at the targeted *cis*-elements. Compared to the original sequence and in combination with a *35Sp::CO* construct, mutations in *S2* and *P1/P2* elements significantly decrease Luc activity in response to CO protein. Consistent with the previous transient expression assay, stable *ft-10* mutant lines transformed with constructs consisting of the 8.1 kb *FT* promoter mutated at *P1/P2* and *S2* elements (*8.1kb-FTp-P1/P2mut::FTcDNA* and *8.1kb-FTp-S2mut::FTcDNA*) flowered significantly later than the original *8.1kb-FTp-FTcDNA ft-10* control. These results indicate that the *S2* and *P1/P2* elements are involved in the photoperiodic induction of *FT* (Adrian et al. 2010). Simultaneously, two CREs named *CO*-

responsive element 1 and 2 (*CORE1* and *CORE2*) were identified within *Block A* with the conserved motif TGTG(N)2-3AT(G) (Tiwari et al. 2010). The *CORE1* element overlaps with the *S2* element previously described. Mutations of *CORE1* and *CORE2* elements fused to the minimal *CaMV/TATA* promoter impair the ability to drive *GUS* expression compared to the original *CORE1* and *CORE2* elements. However, in different constructs consisting of *FT* genomic sequence from -6.5 kb through to ~1.5 kb of the 3' UTR region, called hereafter *pFT::FT*, with mutations in either *CORE* elements, only the mutated version of *CORE2* (*pΔCORE2::FT*) reduced the ability to rescue the *ft-10* mutant phenotype. Additionally, transgenic complementation lines transformed with constructs mutated at both *CORE* elements (*pΔCORE1/2::FT*) flowered significantly later than with *pΔCORE2::FT*. Taken together, these results revealed the role of the *CORE2* element during the photoperiodic induction of *FT* expression, but contrast the impact of the *S2* element previously described and highlight the limitations of complementation experiment analysis. Consistent with the flowering behavior of the *pΔCORE2::FT/ft-10* complementation lines, a combination of electrophoresis mobility shift assay (EMSA) and DNA pull-down approaches indicates that CO binds to *CORE2* *in vitro* (Tiwari et al. 2010). More recently, a ChIP indicates that CO bind to the proximal promoter of *FT* *in vivo* (Song et al. 2012a). Taken together, although the direct binding of CO to the proximal promoter of *FT* is still a matter of debate, the data currently available demonstrate the ability of the CO/NF-Y trimeric complex to mediate *FT* expression through *CORE* elements located at *Block A*.

In *Block C*, our group previously characterized four short conserved “shadow” sequences annotated as *Shadow 1* to 4 (*S1* to *S4*). In addition, prediction of TFBSs revealed an *I*-box, with the consensus sequence (*GATAA*), reported to be implicated in light-regulated genes and a *REalpha* motif, with the consensus sequence (*AACCAA*), known to be associated with gene regulation by phytochromes. A *CCAAT*-box, reported to be involved in gene expression regulation through NF-Y complexes in eukaryotes, has also been identified (Romier et al. 2003). In *Arabidopsis*, an *FT* promoter fragment, located upstream of *Block A* and containing a *CCAAT*-box, has been shown to have the ability to recruit NF-Y complexes *in vitro*, although this interaction seems to be facultative *in planta* (Kumimoto et al. 2008; Tiwari et al. 2010). Additionally, transgenic complementation lines transformed with a mutated version of the *CCAAT*-box in the background of the *pFT:FT* construct (*pFTΔCCAAT::FT*) flowered significantly later than wild-type plants. Furthermore, the

affinity of the NF-Y trimeric complexes to the *CCAAT*-box located at *Block C* has been tested *in vivo* and *in vitro*. A ChIP performed with stable transgenic lines expressing *NF-YB2*, an *NF-Y* homolog, under the control of the *CaMv* promoter, suggests an enrichment of *NF-YB2* to the *CCAAT*-box located at *Block C in vivo* (Cao et al. 2014). Moreover, an EMSA indicates that the *NF-Y* trimeric complex binding is *CCAAT*-box specific *in vitro* since unlabeled oligonucleotides with a mutated version of the *CCAAT*-box lost the ability to compete for the *NF-Y* trimeric complex binding compared to unlabeled oligonucleotides consisting of the original *CCAAT*-box version (Cao et al. 2014).

Additionally, in a complementation assay, each shadow independently mutagenized in the background of the *5.7kbFTp::FTcDNA* construct significantly affects its ability to complement the *ft-10* mutant phenotype (Liu et al, unpublished data).

Recently, our group showed that methylation transgenic plants carrying IR targeting *Block E* showed a reduced the level of *FT* mRNA and a late flowering phenotype compared to wild plants in photoperiod inductive conditions. Taken together, the data defined *Block E* as a novel enhancer involved in the photoperiodic induction of *FT* expression. Phylogenetic footprinting at *Block E* revealed short sequences highly conserved among the Brassicaceae. These sequences are likely to be TFBSs under selective constraints. A putative *CCAAT*-box and an *I*-box (Johan Zicola, Thesis). However, further investigations are required to define the TFBSs required for the photoperiodic induction of *FT* mediated by *Block E*.

Taken together, the data indicate that *Blocks A, C* and *E* represent the photoperiod control regions of *FT*. The data currently available demonstrate that the CO/NF-Y trimeric complex requires *Blocks A* and *C*. However, further investigations are required to define the transcription factors involved in the photoperiodic induction of *FT* mediated by *Block E*. The main challenge remain to characterize how *Blocks C* and *E* regulate *FT* expression. A recent chromatin conformation capture (3C) assay reports an interaction between *Blocks A* and *C* and therefore suggests the formation of a chromatin loop between the distal enhancer and the proximal promoter of *FT* (Cao et al. 2014). Another 3C assay done in our lab rather indicates an interaction of a new control region, called *Block ID*, with the photoperiod control regions, *Blocks A* and *C*. (Liu et al. 2014a). Taken together, the 3C approach done in Arabidopsis at least at the *FT* promoter gave two different points of interaction and brought to question whether this technique, elaborated for long range distances (>1Mb), is appropriate for genomes with small intergenic sequences (Naumova et al. 2012; Bratzel and Turck 2015). The evolution of 3C-based methods and the

emergence of CRISPR/Cas9-derived approaches like the engineered DNA-binding molecule ChIP (enChIP) represent good alternatives to study enhancer-promoter interaction. (Fujita and Fujii 2013; Sati and Cavalli 2017).

1.2.2.2 *Cis*-regulatory elements at *FT* involved in the ambient temperature-mediated response

FT is crucial for the floral transition in response to ambient temperatures (Balasubramanian et al. 2006). Indeed, a null allele of *FT* strongly affects the floral transition at ambient temperatures under SDs (Balasubramanian et al. 2006; Kumar and Wigge 2010; Fernández et al. 2016). PIF4 has been shown to contribute to the thermosensory induction of *FT* expression. ChIP experiments indicate that PIF4 bind to the proximal promoter of *FT* in a temperature-dependent manner. Interestingly, this *FT* genomic region contains a phylogenetically conserved *E*-box (CAAGTGG) located in *Block A*, (Adrian et al. 2010; Kumar et al. 2012). Additionally, increase in temperature has been reported to reduce the level of H2A.Z nucleosomes at *Block A* and may allow the binding of PIF4 to the proximal promoter of *FT* (Kumar and Wigge 2010; Kumar et al. 2012). The thermosensory induction of *FT* expression has also been reported to require CO, suggesting that the CREs involved in the photoperiodic induction of *FT* expression are still active in these conditions (Fernández et al. 2016).

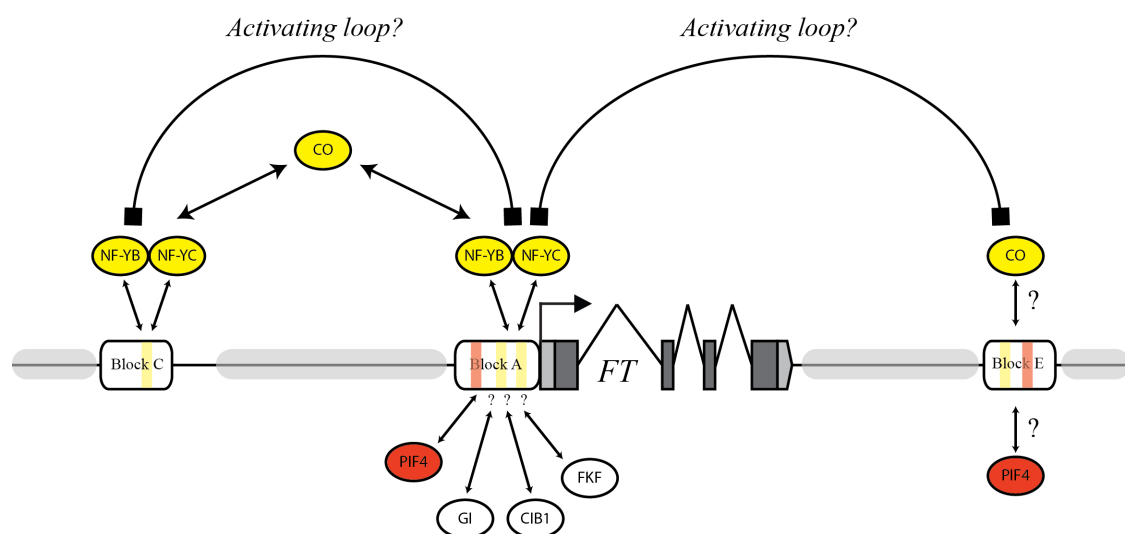


Figure 1: Induction of *FT* expression in the rosette leaves.
See following page

Figure 1: Induction of *FT* expression in the rosette leaves.

The 5,7 kb region upstream of the TSS of *FT* is required for the photoperiodic induction of *FT* expression. It contains two photoperiod control regions, *Block A* and *Block C*, bound by the NF-Y/CO trimeric complex during the floral transition. The downstream genomic region of *FT* carries *Block E* which is also involved in the photoperiodic induction of *FT* expression. In ambient temperature, the induction of *FT* expression requires PIF4, and CO. Putative and functional TFBSs bound by the CO/NF-Y trimeric complex and PIF4 are annotated in yellow and red, respectively. Not mentioned in the text are GI, CIB1, and FKF1 which have been reported to bind the proximal promoter of *FT* (Sawa et al. 2007; Sawa and Kay 2011; Song et al. 2012b). Exons and UTR regions of *FT* are highlighted in dark and light gray, respectively. The sense of the transcription is indicated by an arrow. Gray blocks represent a closed chromatin context.

1.2.3 Repression of *FT* expression

The two MADS-domain transcription factors FLC and SVP have been shown to repress *FT* expression in the rosette leaves. This repression has been reported to be dependent on *CarG*-box CREs (*CC[A/T]6GG*). ChIP experiments indicate that FLC and SVP have the ability to bind the *CarG*-boxes present in the first intron and the proximal promoter of *FT* (Searle et al. 2006; Lee et al. 2007). Additionally, the AP2-like transcription factor SMZ has been shown to bind a genomic region downstream of *FT* (Mathieu et al. 2009).

1.3 Natural variation of flowering time in Arabidopsis**1.3.1 Arabidopsis as a model to study natural variation**

Arabidopsis became a model for molecular plant biology since it presents many advantages including a small diploid genome, a short life cycle (~8 weeks) and an efficient reproduction through self-fertilization. The most commonly used accession Col-0, probably coming from Poland, has been selected for the generation of transferred DNA (T-DNA) insertion lines, due to its particular early flowering phenotype, and was the first sequenced plant genome. For these reasons, the Col-0 accession is to this day the reference accession (The Arabidopsis Genome Initiative 2000; Koornneef and Meinke 2010). However, using a unique reference genotype biases genetic analyses and genome sequencing.

Introduction

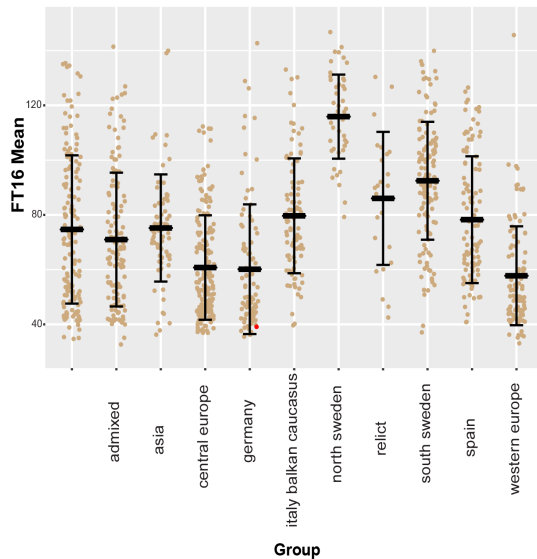
Since a decade, the 1001 genomes consortium made a remarkable effort to establish an atlas of Arabidopsis genetic natural variation (<http://www.1001genomes.org/>). Today, a valuable collection of 1135 Arabidopsis worldwide accession genomes and epigenomes are publicly available (Kwak et al. 2016; The 1001 Genomes Consortium 2016) (http://neomorph.salk.edu/1001_epigenomes.html). This catalog represents a treasurable resource to understand how plants adapt to their environment. Despite an uneven sampling with an overrepresentation in Sweden and Spain for example, these presumably natural inbred lines cover an impressive geographical range from the native Eurasia and Africa to the more recently colonized North America. The population structure revealed eight clusters that broadly correspond to their geographical distribution. Additionally, a group of relicts has been identified mostly in Spain (**Figure 2A**). These relict accessions exhibit an extreme pair-wise divergence and represent survivors of the past glaciation (The 1001 Genomes Consortium 2016).

Arabidopsis accessions show a remarkable range of phenotypic variation including flowering time (The 1001 Genomes Consortium 2016). Indeed, plants grown at 16°C require from 32 days to 145 days to flower (**Figure 2B**). Note that the reference accession Col-0 is one of the earliest accession to flower in these conditions with 38 days and is therefore probably not the best Arabidopsis representative. At 10°C, the earliest accession required 50,7 days to flower whereas the latest needed 157,5 days (**Figure 2C**) (The 1001 Genomes Consortium 2016). The phenotype variation together with the worldwide geographical distribution make Arabidopsis convenient to study the genetic bases of adaptation.

A



B



C

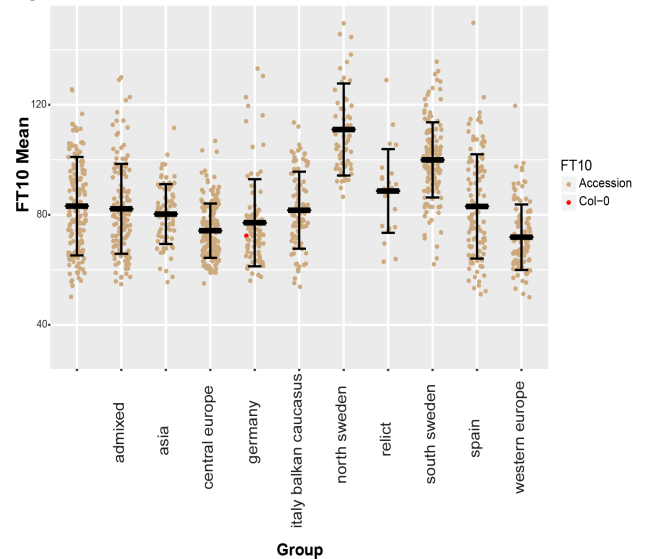


Figure 2: Natural variation of flowering time in the 1001 genomes project.

(A) Picture from the 1001 genomes admixture map where nine genetic groups were identified. Accessions with a genome which derives more than 60% from a given cluster were attributed to the group of the same name as the cluster. Accessions with a genome which do not fit any clusters were attributed to the “admixed” group. Accessions not analyzed with the software package “ADMIXTURE” belong to no group. (B, C) Flowering time of 1163 accessions scored in days until the first flower open with the mean \pm SD for each group. Plants were exposed to LDs at 10°C (B) and 16°C (C) (The 1001 Genomes Consortium 2016). The accessions and the reference strain Col-0 are depicted in brown and red respectively.

1.3.3 Tools to study natural variation in Arabidopsis

1.3.3.1 Identification of the gene

The traditional mutant approach is powerful to characterize genes involved in a particular trait but presents several limitations including the small number of genetic backgrounds of T-DNA.

One of the best examples of genes that cannot be characterized with this approach is *FRI*. Indeed the reference accession Col-0 carries a natural nonfunctional *FRI* allele. For that reason, the natural genetic variations represent a good alternative to discover genes involved in a particular trait. Moreover, this approach allows the characterization of critical ecological genes that contribute to adaptation of plant populations. Different tools were elaborated to identify genomic regions that account for natural phenotypic variations of a particular trait. The traditional linkage mapping and the more recently developed association mapping, based on forward genetics, are the most robust approaches to identify genes. In both methods, the genomic resolution relies on the genetic diversity of the studied population. Linkage and association mappings have limitations but present complementary advantages. The combination of both approaches remains the most powerful way to identify the genomic region underlying natural phenotypic variation.

Traditional linkage mapping, also known as quantitative trait loci (QTL) mapping, uses experimental populations ranging from an F2 population to most elaborated multiparent advanced generation intercross (MAGIC) lines. The most straightforward way to identify genes is to analyze an F2 population but presents many disadvantages. F2 populations derive from biparental crosses and are therefore limited in allelic diversity and genomic resolution. Furthermore, in F2 or later generations, the population is composed of uneven progenies that require genotyping for each phenotyping experiment. For that reason, recombinant inbred lines (RILs), which derive from single seeds descent of F2 individuals, are commonly used. Indeed, a RIL population is nearly homozygous and is therefore considered as immortal. RILs allow replication of phenotyping experiments for different traits in various environmental conditions. Today, more than 60 RIL populations are available (<http://www7.inra.fr/vast/RILs.htm>). Although RILs have a higher number of recombination events than an F2 population, the genomic resolution remain coarse.

Moreover, because the lines in a RIL population have a different genetic background, this approach does not allow the detection of small effect QTLs and epistatic interactions. Various alternative resources were produced to fix this issue. Heterogeneous inbred families (HIFs) are often used to increase the detection for small effect QTL. This population derives from a single recombinant inbred line where only the genomic region of interest segregates for the two parental alleles. However, the individuals of a HIF population have a different background and affect the detection of epistatic interactions. Near-isogenic lines (NILs), also known as introgression lines (ILs), are a good alternative for the detection of both small effect QTLs and epistatic interactions. NILs carry a small introgressed fragment from one parent in a background composed of the other parent. This approach has, however, a lower genomic resolution and presents the disadvantage of being time-consuming. The most advanced experimental populations, the MAGIC lines, use multiple parents to increase the genetic diversity and therefore rise a resolution up to 300 kb (~50 genes) (Koornneef et al. 2004; Bergelson and Roux 2010; Weigel 2012).

Association mapping, also known as genome-wide association studies (GWAS), takes advantage of the genetic diversity of natural populations. This approach is based on the detection of nonrandom allelic associations in a natural population, also known as linkage disequilibrium (LD), to associate genetic variation with phenotypic differences. In contrast to experimental populations, natural populations accumulate recombination events that occurred in long evolutionary period. Therefore, GWAS provides a higher genomic resolution compared to QTL mapping. However, this method is restricted by the sample size, defined by the allele frequency in the GWAS population, and the effect size, related to the phenotypic variation between different alleles.

Higher LD could be explained by different selective pressures but also by population structure. Indeed, the individuals in a population are not equally distantly related to each other. Therefore GWAS are subjected to population structure effects resulting in a high rate of false positive. Different strategies have been developed to correct for population structure and discard false positives (Klasen et al. 2016). As a result of the correction for population structure, the detection by GWAS of “rare” alleles with minor allele frequencies (MAF) below 0,5% remain challenging.

1.3.3.2 Identification of the causal natural variation

Once a genomic region underlying natural variation of a particular trait has been identified, the characterization of the causal polymorphism remains a major challenge. Most of the referenced natural variants that account for phenotypic variation are based on coding sequence analyses.

Indeed, due to the genetic code, the identification of natural variation that changes the protein sequence is straightforward. Single nucleotide polymorphism (SNP) that alter the amino acid sequence, also called non-synonymous SNP (nsSNP), can lead to a premature stop-codon (nonsense mutation), a loss of the normal stop-codon (non-stop mutation) or an amino-acid substitution (missense mutation). Insertions and deletions (InDels) can lead to frameshift mutations. In contrast, mutations that alter gene expression are difficult to detect. The annotation of TFBSs remains coarse in plants in comparison to the ENCODE project in human (Consortium 2012). However, variation in functional CREs allows flexibility in gene expression and consequently phenotype while maintaining the protein function. Despite the fact that functional CREs are challenging to detect, several studies report natural variants of CREs involved in adaptive evolution and suggest that this effect is underestimated.

Next generation sequencing technologies based on paired-end short reads provide a high-resolution map of SNPs. However, the resolution of structural polymorphisms remains dramatically low. Indeed, most of the sequences available result from the alignment of short reads to the Arabidopsis reference genome Col-0. This approach allows the detection of SNPs and short InDels, but larger structural variations encompassing duplications, inversions and TE insertions remain undetectable. *De novo* assemblies, based on the de Bruijn graph or the Eulerian path approaches, are probably the most successful methods to fix this issue but are limited by the presence of repetitive elements (Pevzner et al. 2001; Iqbal et al. 2012). The development of long read sequencing represents a good alternative but remains expensive (Rhoads and Au 2015). The emergence of linked-read sequencing, based on short-read technologies, represents the most promising way but is not extensively studied in plants (Jiao and Schneeberger 2017). For these reasons, considerable efforts were made by the genomic community to provide tools like Pindel or GATK to identify structural variants from re-sequenced genomes (Ye et al. 2009; DePristo et al. 2011). These

algorithms use different approaches including reads-depth, read-pairs and split-reads to recognize the signatures and patterns of structural variations (**Figure 3**) (Tattini et al. 2015).

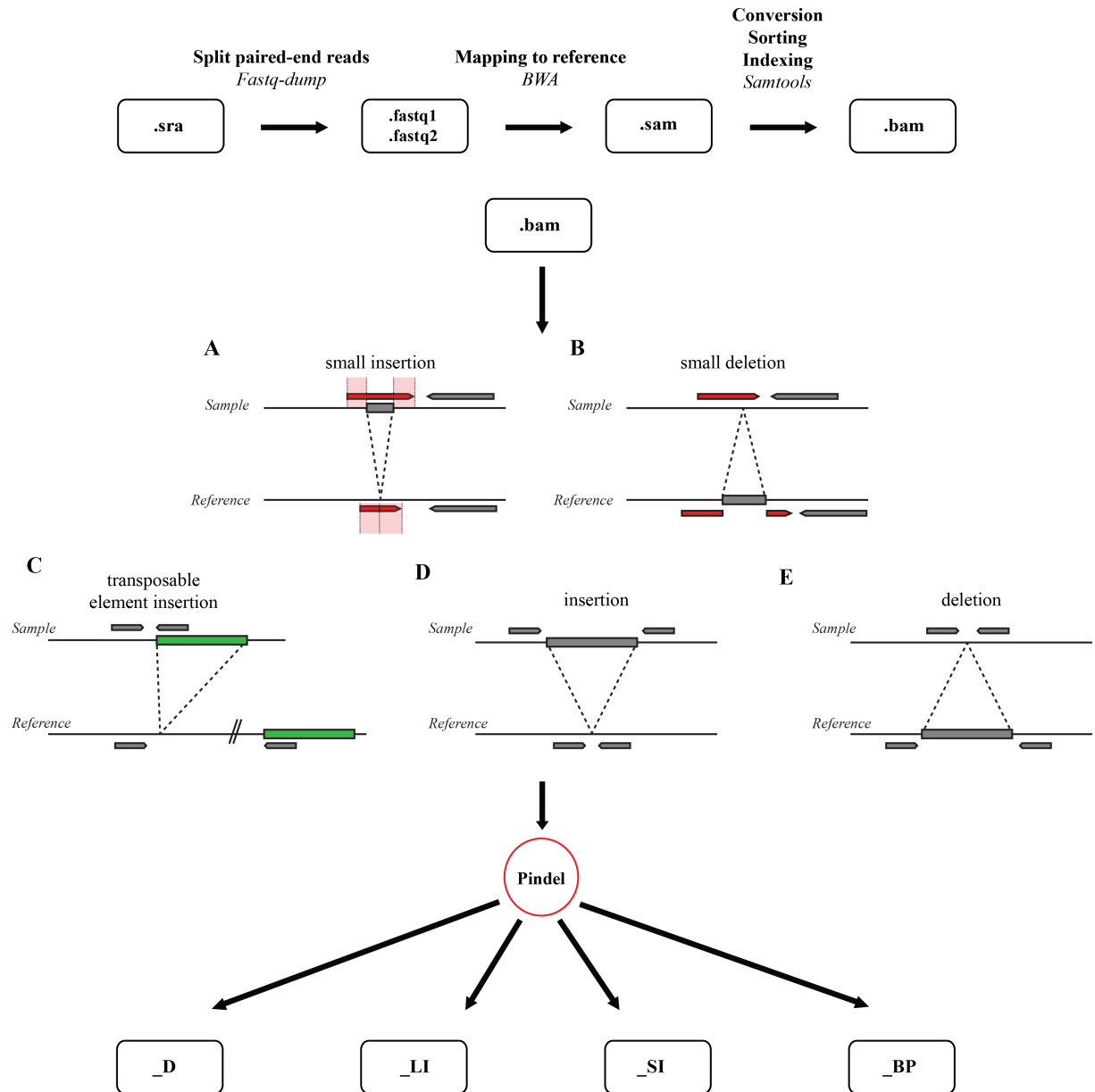


Figure 3: Pindel pipeline to detect structural variants using paired-end short reads. The Pindel uses split-read (A, B) and discordant read-pair events (C, D, E) to detect structural variations. The gray and red arrows correspond to mapped and unmapped reads respectively. The grey rectangles represent InDels. The TEs are annotated with a green rectangle. The Pindel output **_D**, **_LI**, **_SI** and **_BP** correspond to deletion, large insertion, short insertion and discordant read pairs. In C, the double slash on the reference allele indicates that one of the read-pair maps to another locus or chromosomal region.

Additionally, the development of phenotype databases like Arapheno (Seren et al. 2017) allows complementary association analyses to test putative causal natural variations. The ‘gold standard’ approach to testing the functionality of genetic variants relies on complementation experiments. Ultimately, the emergence of genome editing technologies like the CRISPR-Cas9 (clustered regularly interspaced short palindromic repeats/CRISPR-associated protein 9) system will allow the direct proof of causal natural variation and further mechanistic analyses in the regulation of gene expression (Hyun et al. 2015).

1.3.4 Natural genetic variation affecting flowering time in Arabidopsis

Due to their sessile nature, plants are perpetually exposed to seasonal changes in photoperiod and temperature, and have developed different strategies to flower in the optimal conditions and therefore achieve reproductive success.

The late flowering character has been suggested to be ancestral in plants (Johanson et al. 2000; Roux et al. 2006). Indeed, the behavior of single season Arabidopsis accessions has been demonstrated to arise through the generation of weak alleles of floral repressor genes like *FLC* and *FRI* (Michaels et al. 2003). The main causal natural variations at *FRI* have been identified. They affect both the transcriptional and protein level due to a deletion of 16 bp leading to a frame shift in the *FRI-Col-0* allele and a deletion of 376 bp present in the *FRI-Ler* allele, respectively. Some additional SNPs leading to nonsense mutations have also been characterized (Shindo et al. 2005). *FLC* is a common flowering time QTL in both laboratory and field condition experiments (Salomé et al. 2011; Grillo et al. 2013; Dittmar et al. 2014; The 1001 Genomes Consortium 2016). Until now, the causal natural variations correspond exclusively to structural variations in the first intron of *FLC* that interfere with regulatory regions. *Ler*, for example, has a weak *FLC* allele due to a *Mutator*-like element (MULE) TE inserted in the first intron (Gazzani et al. 2003). Recent efforts have been made with a set of 1307 accessions to characterize different *FLC* haplotypes based on noncoding sequence variation that underlie *FLC* expression level and epigenetic silencing rate (Li et al. 2014). A structural variant of *FLM* has been characterized to modulate flowering time in a significant number of Arabidopsis accessions. This *FLM* haplotype contains a LINE retrotransposon that modifies the architecture of the gene and accelerates the floral transition in cool temperatures (Lutz et al. 2015). Additionally, *VIN3* and *SVP* often

colocalize with flowering time QTLs (Grillo et al. 2013; Agren et al. 2017), however, the causal genetic variation remain poorly described.

Most of the flowering time QTLs identified until now correspond to floral repressing genes more likely to be the targets of selection for early flowering. The identification of flowering time QTLs associated with natural variation at floral promoting genes remains exceptional compared to floral repressing genes. Among the flowering time QTLs mapped to floral promoting genes is *CRY2*. A single amino acid substitution at *CRY2* has been shown to reduce its light induced downregulation in SDs, leading ultimately to early flowering (El-Din El-Assal et al. 2001; El-Din El-Assal et al. 2003). Additionally, variation at *TSF* has been associated with differences in flowering time in field experiments (Brachi et al. 2013). Interestingly, a GWAS done on a global population identified *DELAY OF GERMINATION (DOG1)*, a gene known to have a major role on seed dormancy, as a flowering time QTL (Bentsink et al. 2006; The 1001 Genomes Consortium 2016). Interestingly, pleiotropy between flowering time and seed germination has also been reported for *FLC* (Chiang et al. 2009).

1.3.5 Difference of flowering time due to natural variation at *FT*

Natural variation at the *FT* locus has already been reported to be associated with differences in the floral transition (Schwartz et al. 2009; Li et al. 2010; Strange et al. 2011; Brachi et al. 2013; Sanchez-Bermejo and Balasubramanian 2015; The 1001 Genomes Consortium 2016). However, flowering time QTLs closely linked to *FT* remain rare and exclusively mapped to non-coding variation.

Natural variation of *FT* promoter length has been reported to modulate the photoperiodic induction of *FT* expression in the rosette leaves (Liu et al. 2014a). Structural variations at the promoter of *FT* define a long, a medium and a short *FT* promoter type. The medium *FT* promoter type is the most common *FT* haplotype in *Arabidopsis* and has been shown by phylogenetic analysis to represent the ancestral state. The long and the short *FT* promoter types are respectively due to an insertion of ~1.1 kb and a deletion of ~250 bp. Allele-specific expression (ASE) of *FT* in F1 hybrids revealed differences in response to LDs between the three *FT* haplotypes. The effect of the three main structural *FT* variants has

Introduction

been tested on available fitness data from field experiments done with 82 accessions genotyped for the *FT* promoter types. Interestingly, the analysis revealed no significant differences in bolting time but rather in seed production. The long *FT* promoter had been associated with a fitness advantage in overwintering plants (Korves et al. 2007; Liu et al. 2014a).

A flowering time QTL was closely linked to the promoter region of *FT* in Est-1 (ID: 6916), an accession coming from Sweden (Schwartz et al. 2009). This *FT* haplotype has been reported to alter *FT* expression in LDs in a dosage-dependent manner. Indeed, in a population of NIL homozygous either for *FT-Est-1* or *FT-Col-0*, NIL-Est-1 had a lower level of *FT* mRNA compared to NIL-Col-0 whereas the heterozygous one has an intermediate phenotype.

In Ull2-5 (ID: 6974), another accession coming from Sweden, a flowering time QTL was also mapped to the promoter sequence of *FT*. A SNP in the functional *CCAAT*-box located in *Block C* has been proposed to affect *FT* expression in the rosette leaves in response to LDs (Strange et al. 2011). This *FT* haplotype has been shown to cause a difference in flowering time in an *FLC*-independent manner since the *FT-Ull2-5* allele in an F2 population resulting from the cross between Ull2-5 and the *FLC* mutant *flc-2* delays the floral transition regardless of the *FLC* allele. However, the flowering time QTL mapped to the *FT-Ull2-5* haplotype was interestingly affected by the length of vernalization. Indeed the *FT-Ull2-5* haplotype explained only 43% and 24% of the variance in flowering time after exposure to 8 and 14 weeks of vernalization, respectively. Additionally, Ull2-5 has been reported to have a strong *FLC* haplotype and to therefore show a slow vernalization behavior (Li et al. 2014).

Quite recently, another flowering time QTL was mapped to the promoter region of *FT* in Tsu-0 (ID: 7373) an accession coming from Japan (Sanchez-Bermejo and Balasubramanian 2015). In a HIF population homozygous either for *FT-Tsu-0* or *FT-Col-0*, the *FT-Tsu-0* allele has been reported to affect *FT* expression in LDs. Structural variations and SNPs located ~3 kb upstream of the start codon of *FT* are shared with Ull2-5 have been suggested to explain this differential *FT* expression.

In C24 (ID: 6906), an accession coming from Portugal, *FT* was also associated with a difference of flowering time in a RIL population (Huang et al. 2013). Natural variation in

the promoter sequence of *FT-C24* has been demonstrated to significantly alter *FT* expression in F1 hybrids and in a HIF population. The C24 accession has a short *FT* promoter type, already described to reduce *FT* expression in LDs compare to the medium and long *FT* promoter types (Liu et al. 2014a).

In a GWAS for flowering time done in the field, *FT* was detected for a difference in bolting time and interval between bolting and flowering time (Brachi et al. 2013). However, surprisingly, *TSF* appears to be the central floral integrator gene in this field experiment. This finding is quite unexpected knowing the weak phenotype of a null allele of *TSF* in a Col-0 background (Yamaguchi et al. 2005).

Recently, the 1001 genomes consortium identified *FT* in a GWAS for flowering time done in LD at 10°C (The 1001 Genomes Consortium 2016). Interestingly, the peak of association at *FT* was not significant at 16°C.

2. Aim of the study

FT has been shown to be essential for the floral transition in response to photoperiod and ambient temperature in single season accessions. Previous studies reported natural genetic variation at *FT* underlying flowering time differences.

The aim of the present study was to evaluate whether *FT* could be a target of selection to modulate the floral transition in *Arabidopsis*. Until now, flowering time QTLs closely linked to *FT* are exclusively mapped to CREs nearby *FT*. For that reason, the genomic sequences available in the 1001 genomes project were investigated, with a focus on the photoperiod control regions of *FT*. *FT* haplotypes containing natural variation in the annotated TFBSs or with TEP were tested in F1 hybrids. The ASE profile of *FT* was characterized for induction and tissue specificity. The 5,7 kb promoter region of the less photoperiod insensitive *FT* haplotypes was tested by complementation with the aim to identify the causal natural variation.

The most photoperiod insensitive *FT* haplotypes turned out to share a SNP in the functional *CCAAT*-box located at *Block C*. Among the three accessions related to these photoperiod insensitive *FT* haplotypes is Ull2-5, already reported for its distinct slow vernalization behavior. Because different selective pressure could drive natural selection at *FT* in this population, the flowering behavior of the candidate accessions exposed to an extensive period of cold temperature was compared to the flowering time of null allelic mutants of *FT*. The contribution of *FT* to the floral transition was evaluated after different periods of cold treatment.

3. Results

3.1 Natural variation in the coding sequence of *FT*

Until now, flowering time QTLs closely linked to *FT* were associated with natural variation noncoding regions. To verify this trend, I first analyzed the sequences available from the 1001 genomes project with the aim to identify structural variations and nsSNPs at the coding sequence of *FT*. In a global population of 1135 accessions, the Pindel pipeline did not detect any structural variations, including InDels and TEs. In contrast, the nucleotide variant calling pipeline characterized a total of 10 nsSNP *FT* variants leading to 8 missense mutations (V18I, F22L, K29N, V30F, Q34R, Q49L, A158T, C164Y) and one non-stop mutation (*176Y) (**Figure 4A**).

Aside the missense *FT* variant leading to the V18I amino acid substitution with a MAF of 1,76%, every *FT* haplotype defined had a MAF below 0,5% and was therefore considered as “rare” (**Figure 4A**). The missense *FT* variants V18I, F22L, V30F, A158T, and C164Y disturb amino acids highly conserved among 88 *FT* homologs of flowering plants (Ho and Weigel 2014) and therefore represent candidates for altered *FT* protein function or structure. However, the V18I and A158T amino acid substitutions are more likely to have no effect on the protein function since both amino acid substitutions contain side chains with low reactivity and are therefore rarely involved in protein function. Moreover, both amino acid substitutions affect amino acids with similar chemical properties. The F22L and V30F amino acid substitutions disturb amino acids with highly reactive side chains and are replaced by amino acids with different chemical properties. To evaluate their effect on the protein structure, I simulated the F22L and V30F amino acid substitutions on the available crystal structure of *FT* (Ahn et al. 2006). F22L and V30F amino acid substitutions show significant van der Waals overlaps and therefore potentially affect the protein structure of *FT* (**Figures 4B and 4C**). The C164Y amino acid substitution concerns amino acids with different chemical properties. Cysteines are frequently involved in disulphide bonds that stabilize the protein structure. Unfortunately, I could not test this amino acid substitution on the crystal structure of *FT* since the Cys164 amino acid was replaced by a Serine for the crystallization. The *176Y amino acid substitution leads potentially to a longer *FT* protein. However, several putative stop-codons are present 9, 14 and 15 codons downstream of the normal stop-codon.

Results

I applied a generalized linear model (GLM) to detect any significant associations between *FT* haplotypes, containing coding-sequence variation, and geographical coordinates. After correction for population structure, the analysis does not reveal particular latitudinal and longitudinal clines (**Figure 4D, Supplementary Figure 5**).

Taken together, natural variation in the coding sequence that can alter FT protein function or structure is present at a low frequency in a global population of 1135 accessions and suggest a high conservation of the protein sequence of FT in Arabidopsis.

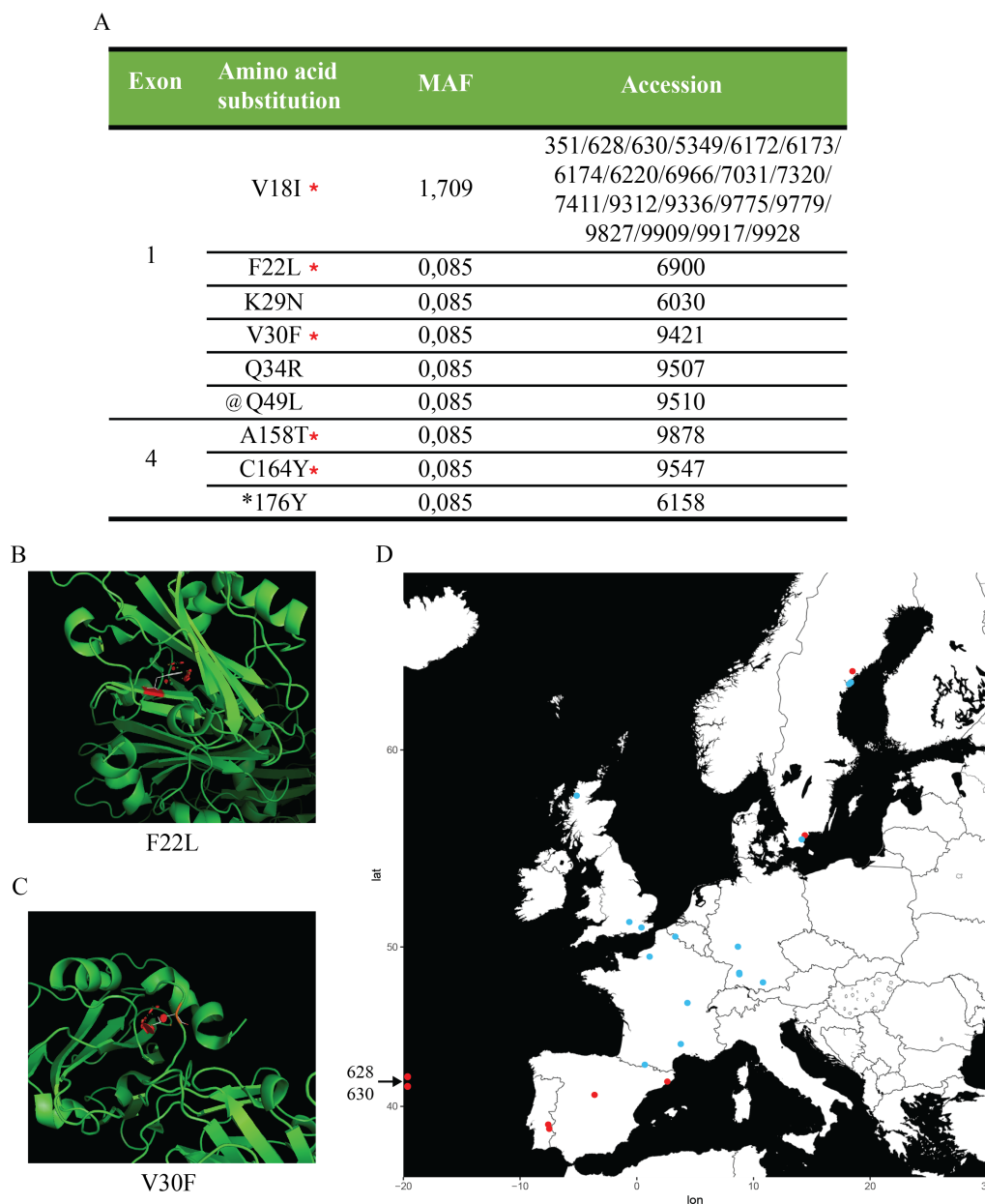


Figure 4: Nonsynonymous SNPs at *FT* in Arabidopsis.
See following page

Figure 4: Nonsynonymous SNPs at *FT* in *Arabidopsis*.

(A) Summary table of accessions with a nsSNP at *FT*. The frequency indicates the percentage of *FT* haplotype in a global population of 1166 accessions. Substitutions of highly conserved amino acid are annotated with a red star (Ho and Weigel 2014). The accession 9910 annotated with @ shows two SNPs within the same codon leading to the Q49L amino acid substitution. The translation stop-codon is indicated with a black star. MAF corresponds to the minor allele frequency in % in a global population of 1135 accessions provided by the 1001 genomes project. (B and C) Ribbon representation of the crystal structure of the FT protein with the F22L and the V30F amino acid substitutions in (B) and (C), respectively (Ahn et al. 2006). Large red disks indicate significant van der Waals overlap. (D) Geographical distribution of accessions containing a nsSNP at *FT* in Europe. Blue dots represent the localization of accessions with the V18I FT variant. Red dots indicate the localization of the candidate accessions. The accessions 628 and 630 come from the USA.

3.2 Natural variation in the non-coding sequence of *FT***3.2.1 Natural variation in the promoter length of *FT***

The geographical distribution of the three *FT* promoter types previously described in a global population of 80 accessions, revealed an accumulation of the long and short *FT* promoter types in Northern and Southern latitudes, respectively. (Cao et al. 2011; Liu et al. 2014a).

The combination of re-sequenced *Arabidopsis* accessions provided by the 1001 genomes project and the bioinformatics tools to detect structural variants from paired-end short reads represents an excellent opportunity to confirm this geographical trend with a more representative set of accessions. In this assay, sequences with reads covering the specific Col-0 structural variations, previously characterized by our group, were defined as long *FT* promoter types. The medium *FT* promoter type was characterized by the absence of reads on the structural variations corresponding to the two Col-0 specific insertions and confirmed by split-read events following the PINDEL pipeline. In addition to the two distinct structural variations present in Col-0, the short *FT* promoter type was identified by the absence of reads at the two Ws-0 specific deletions and confirmed by split-read events following the PINDEL pipeline (Figures 5A and 5B). After filtering for coverage sequencing at the *FT* locus, a total of 598 accessions were analyzed. To estimate the accuracy of the structural variation prediction assay using paired-end short reads, we checked a set of 132 accessions genotyped at the *FT* promoter. The analysis shows a high

Results

accuracy since more than 99% of the accessions genotyped were confirmed by sequence analysis.

I applied a GLM model to detect any significant associations between *FT* promoter types and geographical coordinates. Unexpectedly, the analysis revealed that the medium and long *FT* promoter types are associated with respectively higher and lower latitudes, independently of the population structure (**Figure 5C, Supplementary Figures 3 and 4**). Additionally, the structural variation detection study revealed the presence of the short *FT* promoter type in 5 accessions coming from Sweden (**Figure 5C**). This result was unexpected since this *FT* haplotype has been shown to accumulate in Southern latitudes. However, the Swedish populations have not been tested in this study (Liu et al. 2014a). Interestingly, the relict populations, known as survivors of the last glaciation, contain mainly the long *FT* promoter type (78,9% and 11,1% of long and medium *FT* promoter types, respectively) (**Supplementary Table 5**). This result was surprising since the medium *FT* promoter had been reported by phylogenetic analyses to be the ancestral character in *Arabidopsis*. Indeed, the specific Col-0 insertion, proximal to *FT*, consists mainly of a sequence nearly identical to a region on the Chromosome 5, which suggests a recent event (Liu et al. 2014a).

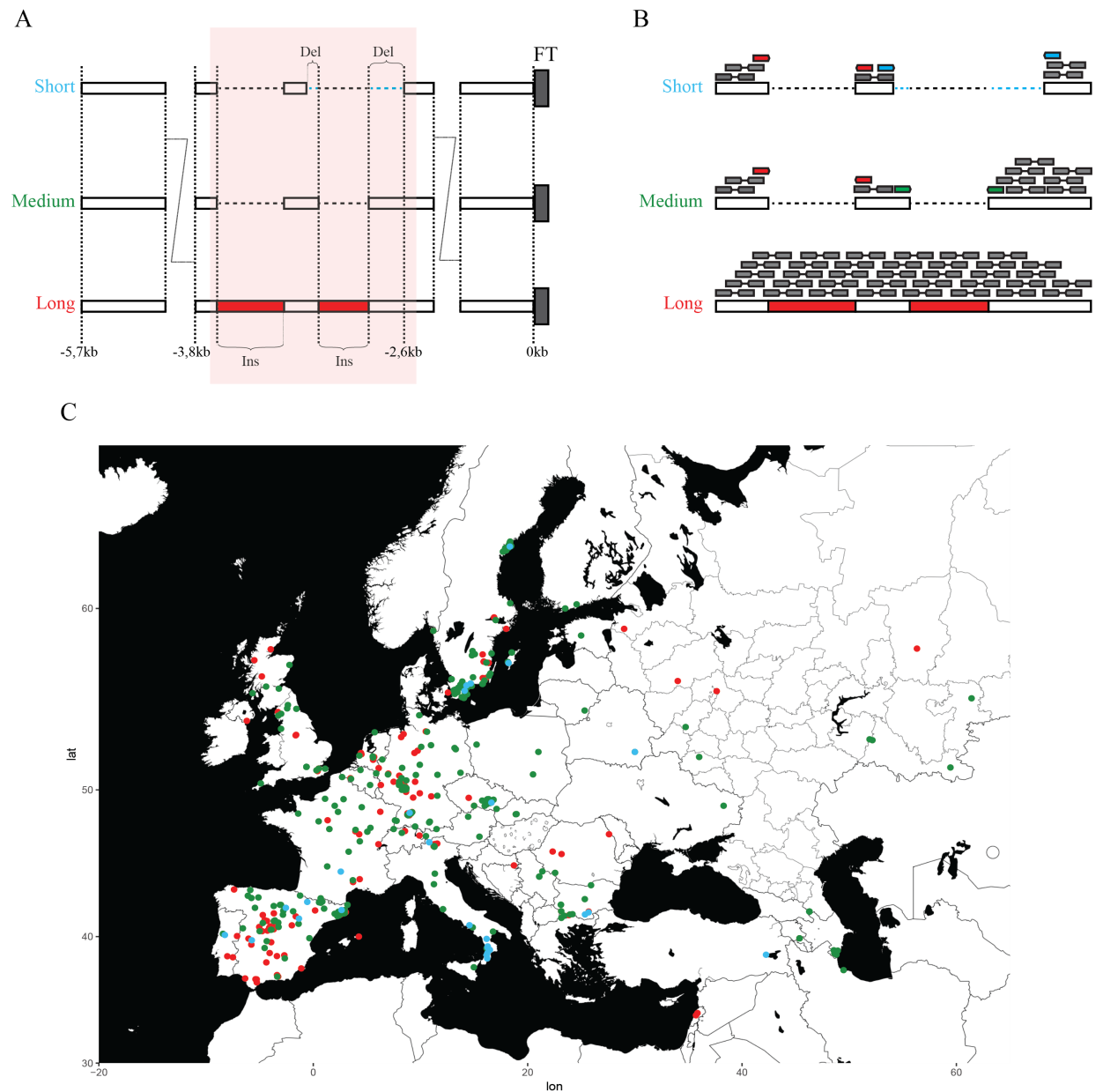


Figure 5: Geographical distribution of the *FT* promoter types.

(A) Schematic representation of the three main *FT* promoter types already described (Liu et al. 2014a). Insertions (Ins) and deletions (Del) are illustrated with red boxes and blue dashed lines, respectively. Promoter length breaks are represented as “zigzag”. (B) Schematic representation of the pipeline used to detect the three main *FT* promoter types using paired-end short reads. Sequences with reads covering the InDels previously described were defined as long *FT* promoter types. Sequences with missing reads and split-reads events at the InDel defining the medium *FT* promoter type were categorized as medium *FT* promoter type. Sequences with missing reads and split-reads events at the InDel defining the short *FT* promoter type were classified as short *FT* promoter type. (C) Geographical distribution of the *FT* promoter types in Europe. The long, medium and short *FT* promoter types are annotated with red, green and blue dots, respectively.

3.2.2 Natural variation in the photoperiod control regions of *FT*

To test the conservation of the *FT* locus in Arabidopsis, we first checked the variant calling format (VCF) available in the 1001 genomes project. VCF files give information about the presence of SNPs and small InDels but miss the annotation of larger structural variations (<https://samtools.github.io/hts-specs/VCFv4.2.pdf>). However, when neither the reference nor the alternative allele is annotated, this indicates the absence of reads and could suggest the presence of InDels that have to be confirmed either by polymerase chain reaction (PCR) or with the bioinformatics tools available to study structural variations from paired-end short reads. As expected, the *FT* locus, including coding and non-coding regions, is highly conserved in Arabidopsis. We can notice the signature of the InDels defining the three main *FT* promoter types. However, the photoperiodic control regions, including *Blocks A*, *C* and *E*, are not more conserved than the rest of the *FT* locus (**Figure 6**).

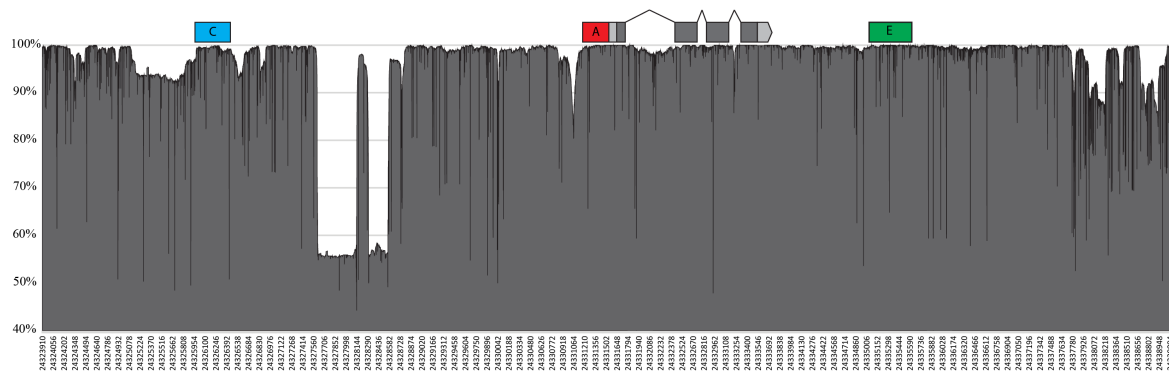


Figure 6: Conservation of *FT* in Arabidopsis.

Conservation of the *FT-Col-0* haplotype among the 1001 genomes project. The y and x axis correspond to the conservation of the *FT-Col-0* haplotype in percentage and the nucleotide position on the Chromosome 1, respectively. *Blocks A*, *C*, and *E* are annotated in red, blue and green, respectively. Exons and UTR regions of *FT* are highlighted in dark and light grays, respectively.

To identify *FT* haplotypes affected in the floral transition, we first applied a phylogenetic footprinting approach using the Brassicaceae family to characterize highly conserved nucleotides within the photoperiod control regions. Consistent with the previous functional characterizations, the motifs of the *CORE* responsive elements and the *CCAAT*-box, located at *Blocks A* and *C*, respectively, are preserved after filtering (**Figure 7 and 8**). Additionally, putative TFBSs previously annotated show a high conservation. These putative TFBSs include the *Shadow 1* and the *E*-box at *Block A*, and the *Shadows 1* to 4 as well as the *REalpha* motif at *Block C*. However, we also remark the presence of highly

conserved motifs, not previously characterized and likely to be TFBSs under selective constraints.

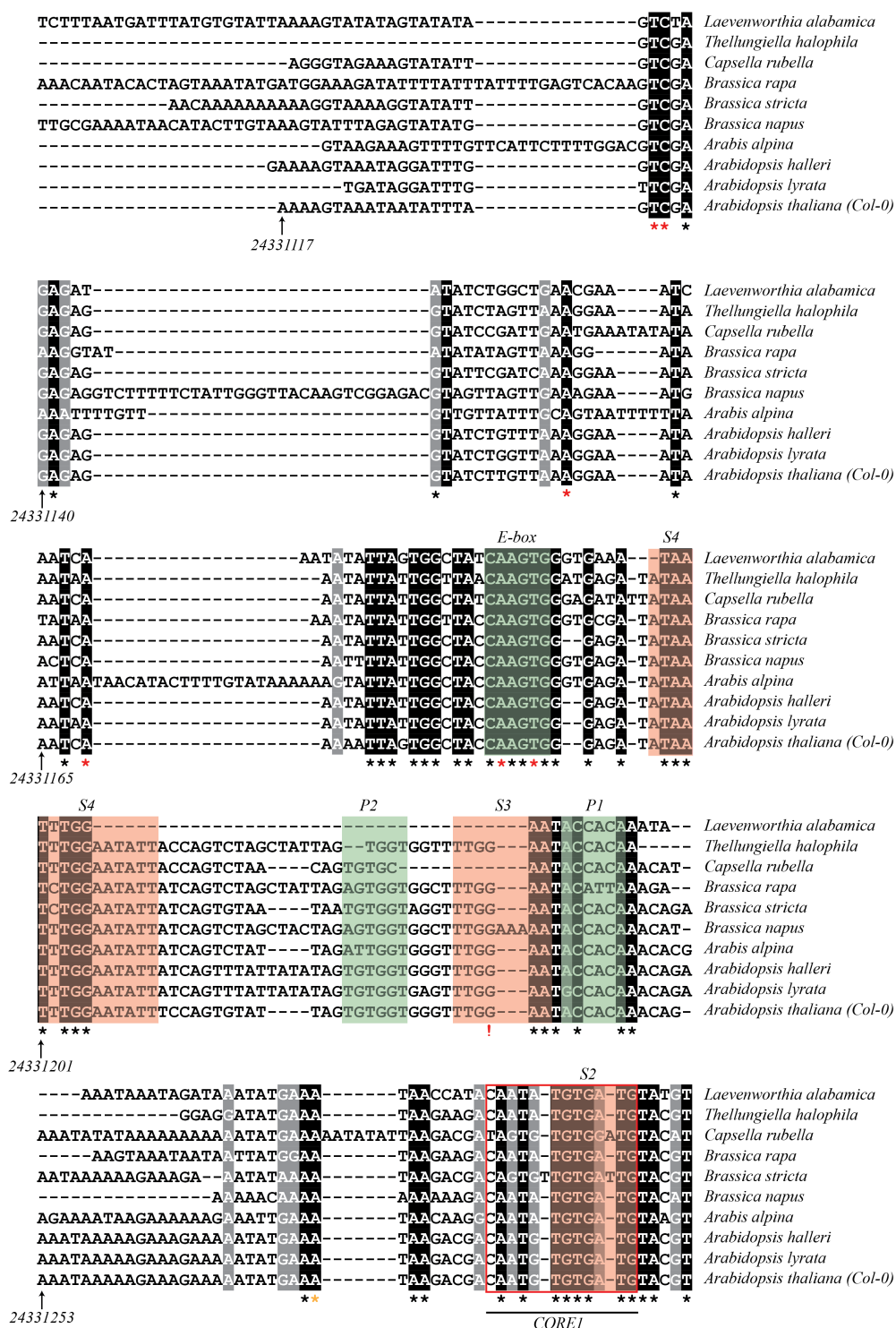


Figure 7: Natural variation in the proximal promoter (Block A) of *FT*.
See following page

S1

AGAATCAGTTT TAGATT C--TAGTACATCACTAGACAAGAAAA GATTGTGGTGATG AGT	<i>Laeneworthia alabamica</i>
AGAATCAGTTT TAGATT C--TCGTACATCACTGGACAAGAAAA GATTGTGGTGATG AGT	<i>Thellungiella halophila</i>
AGAATCAGTTT TAGATT C--TACTACATCAATAGACAAGAAAA GATTGTGGTGATG AGT	<i>Capsella rubella</i>
AGAATCAGTTT TGGATT C--TCGTCCATCACTGTACAGAAAAA GATTGCCGTGATGA GC	<i>Brassica rapa</i>
AGAATCGGTTTT TAGATT C--TAGTACATCAATAGACAAGAAAA GATTGTGGTGATG AGT	<i>Brassica stricta</i>
AGAATCAGTTT TAGATT C--TCGTACATCACTAGACAAGAAATATATT TGTGGTGATG AGT	<i>Brassica napus</i>
AGAATCAGTTT TAGATT C--CCATAGAACACTAGACAAGAAAA GATTGTGGTGATG AGT	<i>Arabis alpina</i>
AGAATCAGTTT TAGATT GC--TAGTACAACTAGACAAGAAAA GATTGTGGTTATG ATT	<i>Arabidopsis halleri</i>
AGAATCAGTTT TAGATT GCATATAGTACATCATATACAAGAAAA GATTGTGGTTATG ATT	<i>Arabidopsis lyrata</i>
AGAATCAGTTT TAGATT C--TAGTACATCAATAGACAAGAAAA GATTGTGGTTATG ATT	<i>Arabidopsis thaliana (Col-0)</i>

* * * * *

↑
24331304

CORE2

TCACC GA -CCC GT -TTAATGTA AAAT CCCT-AAACACTATATATATAT-----GTGTG	<i>Laeneworthia alabamica</i>
TCACC GCAC CCCC GA -TTAGTGC AAT CCCCG-AAACACTATATATATATATAAGTTAC ACG	<i>Thellungiella halophila</i>
TCACCA AA -CCCA GT TTAATGCA AAAT CTTGAAAACAGTATAAAATATAT-----G	<i>Capsella rubella</i>
TCTCC GA -CCC GA G-TTA AA ACC ACT CCCA-AGACACTATATATATAC---GTGACATG	<i>Brassica rapa</i>
TCACC GA -CCC GA G-TTAATGCA AAAT CCCG-AAACAGTATAAAATTTAT-----G	<i>Brassica stricta</i>
TCACC GA -CCC GA G-CTAATTC AAT CC-----G	<i>Brassica napus</i>
TCACC GA -CCC GA G-TTAATGCA AAAT CCCG-AAACACTATAAATATAA-----G	<i>Arabis alpina</i>
TCACC GA -CCC GA G-TTAATGCA AAAT -CCG-AAACAGTATAAAATATGT-----G	<i>Arabidopsis halleri</i>
TCACC GA -CCC GA G-TTAATGCA AAAT -CCG-AAACAGTATAAAATATGT-----G	<i>Arabidopsis lyrata</i>
TCACC GA -CCC GA G-TTAATGCA AAAT -CCG-AAACAGTATAAAATATGT-----G	<i>Arabidopsis thaliana (Col-0)</i>

** * * *

↑
24331362

Multiple alignments of *Block A* from different Brassicaceae species. The functional *CORE1* and *CORE2* elements are surrounded by a red square. The putative *E-box* TFBS and the palindromic sequences *P1/P2* are annotated in green. The previously identified conserved stretches named *Shadow 1* to *4* are highlighted in red. The multiple alignments were done with MUSCLE (Edgar 2004). The shadings were done with BOXSHADE, and correspond to the degree of conservation. Nucleotides conserved among species and the 1001 genomes project are annotated with a black asterisk. SNPs from the 1001 genomes project in nucleotides conserved among the Brassicaceae species tested are annotated with a red or yellow asterisk depending on the number of accessions sharing the SNP (< 10 and > 10 respectively). The position on the Chromosome 1 is indicated by an arrow.

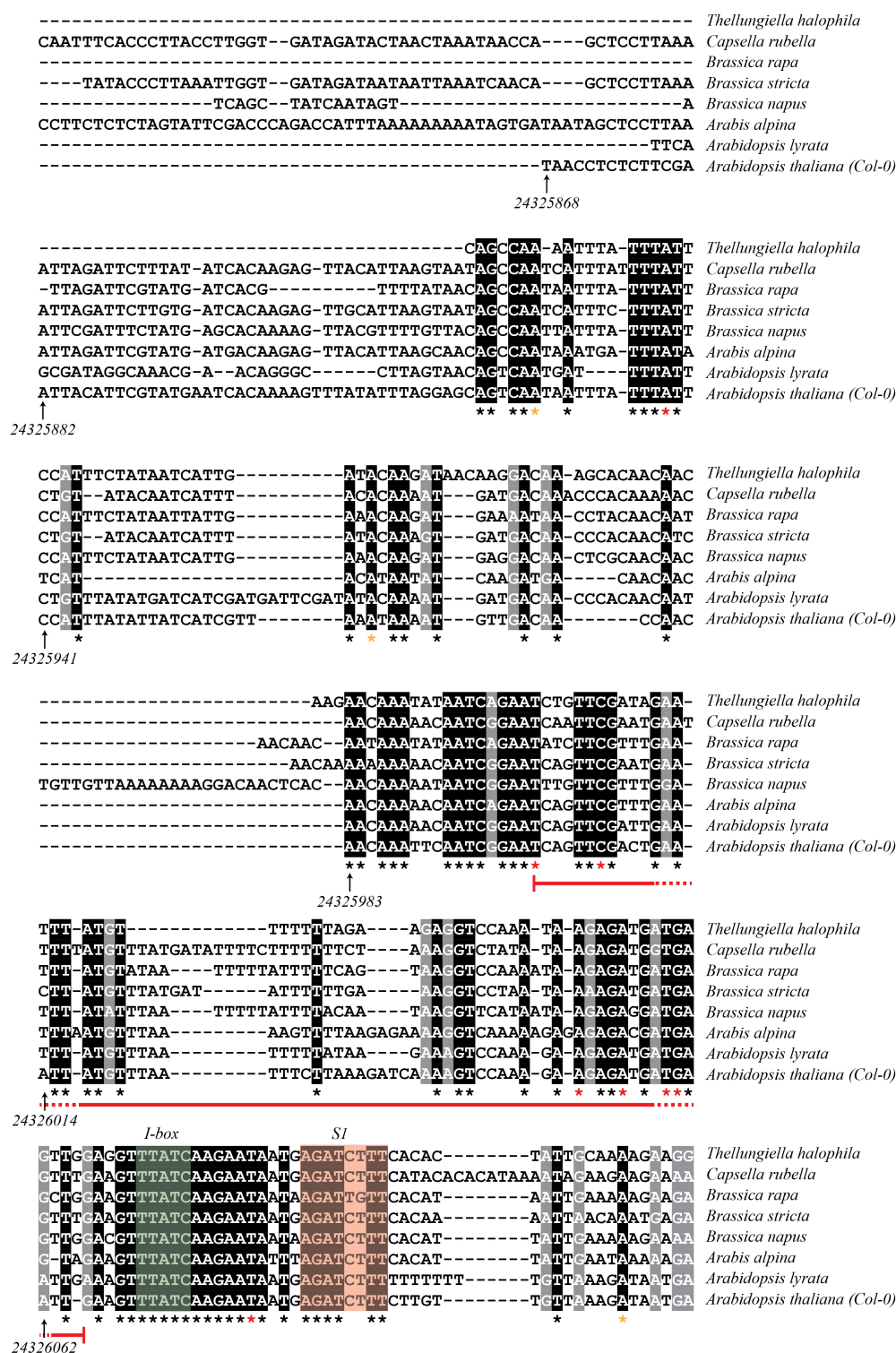


Figure 8: Natural variation in the distal enhancer (Block C) of *FT*.

See following page

Results

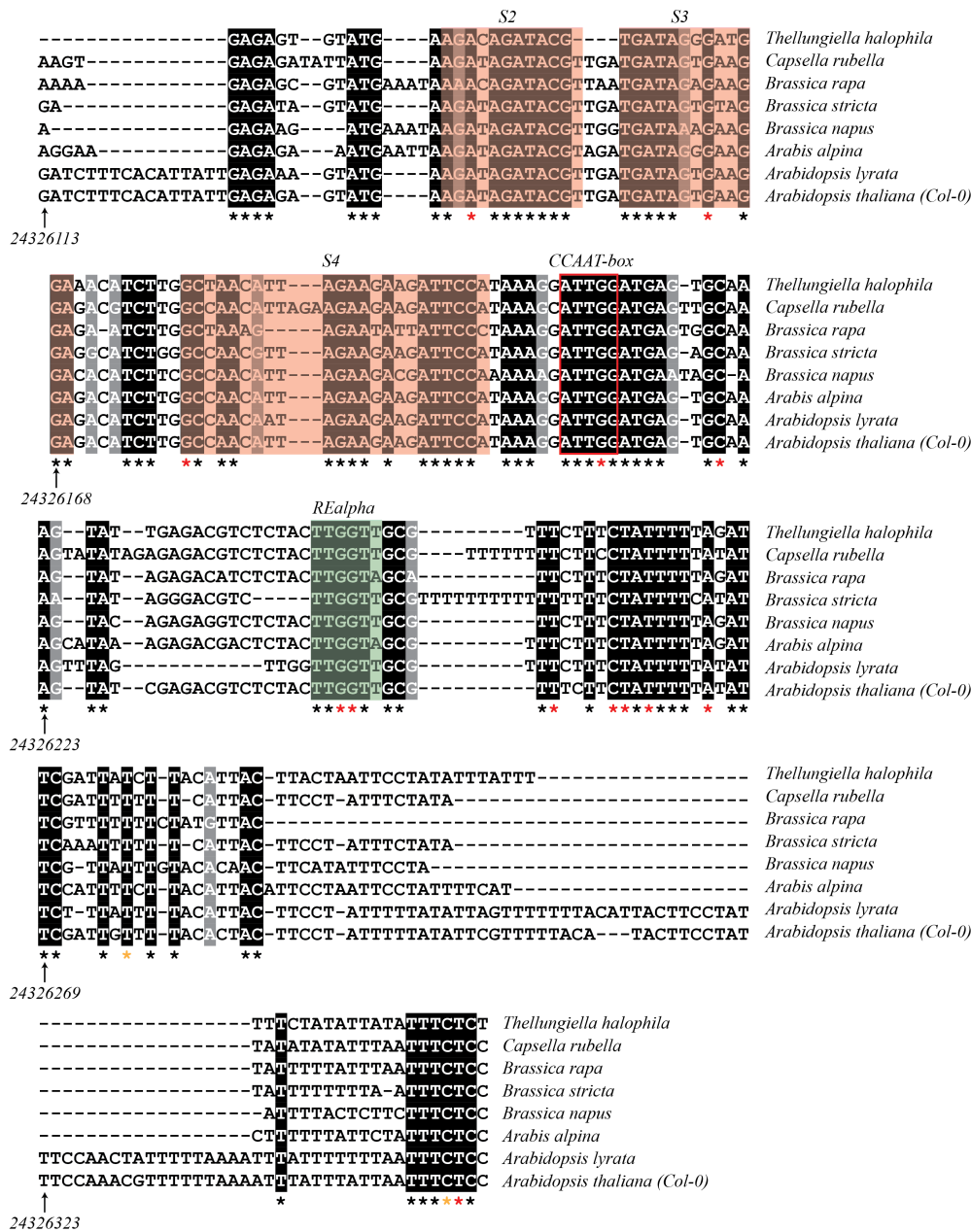


Figure 8: Natural variation in the distal enhancer (*Block C*) of *FT*.

Multiple alignments of *Block C* from different Brassicaceae species. The functional CCAAT-box is surrounded by a red square. The putative REalpha and I-box TFBS are annotated in green. The previously identified conserved stretches named *Shadow 1* to 4 are highlighted in red. The multiple alignments were done with MUSCLE (Edgar 2004). The shadings were done with BOXSHADE, and correspond to the degree of conservation. Nucleotides conserved among species and the 1001 genomes project are annotated with a black asterisk. SNPs from the 1001 genomes project in nucleotides conserved among the Brassicaceae species tested are annotated with a red or yellow asterisk depending on the number of accessions sharing the SNP (< 10 and > 10 respectively). The red line corresponds to the deletion present in the accession 7323. The position on the Chromosome 1 is indicated by an arrow.

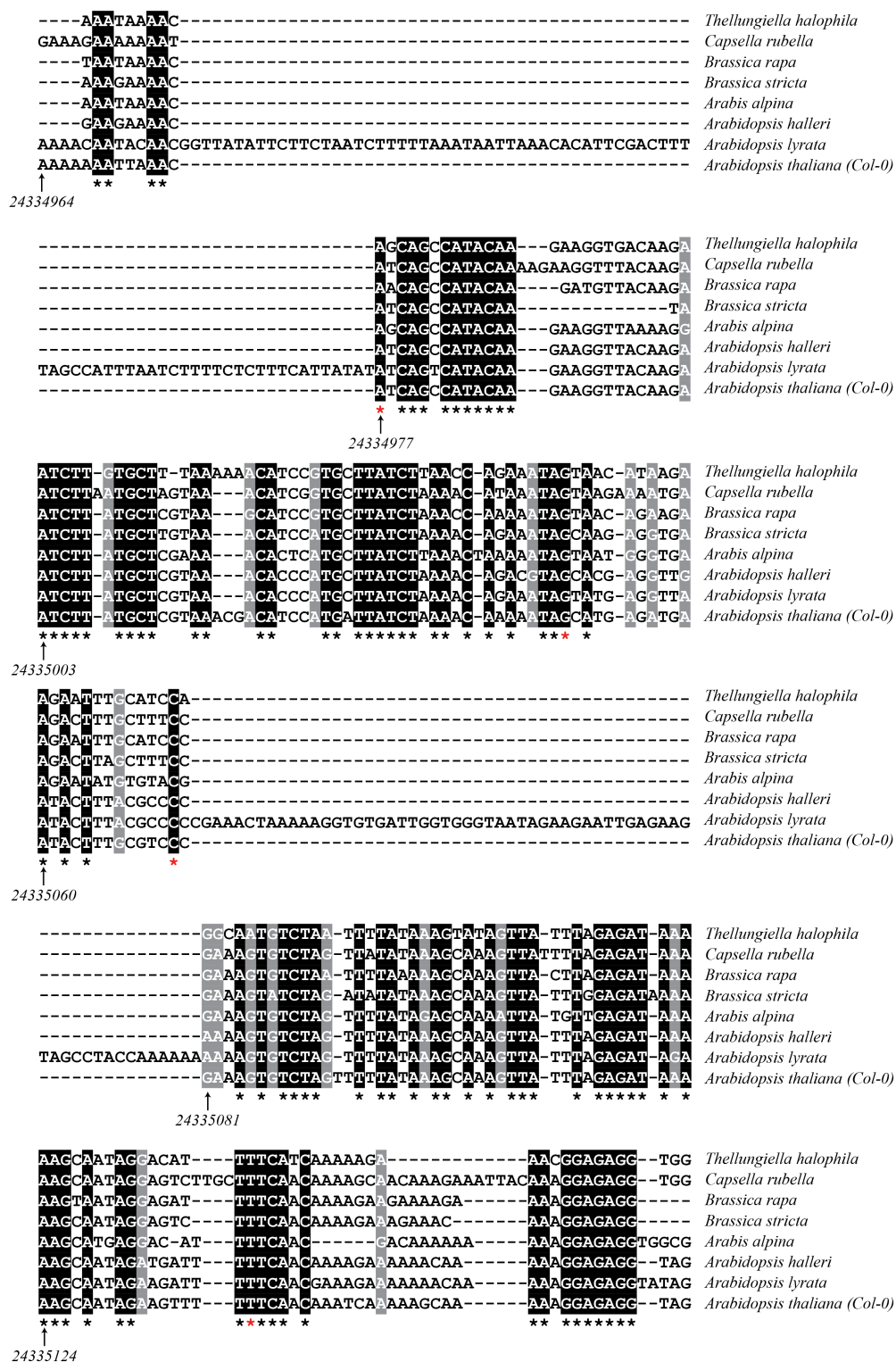


Figure 9: Natural variation in the novel enhancer (Block E) of *FT*.
See following page

Results

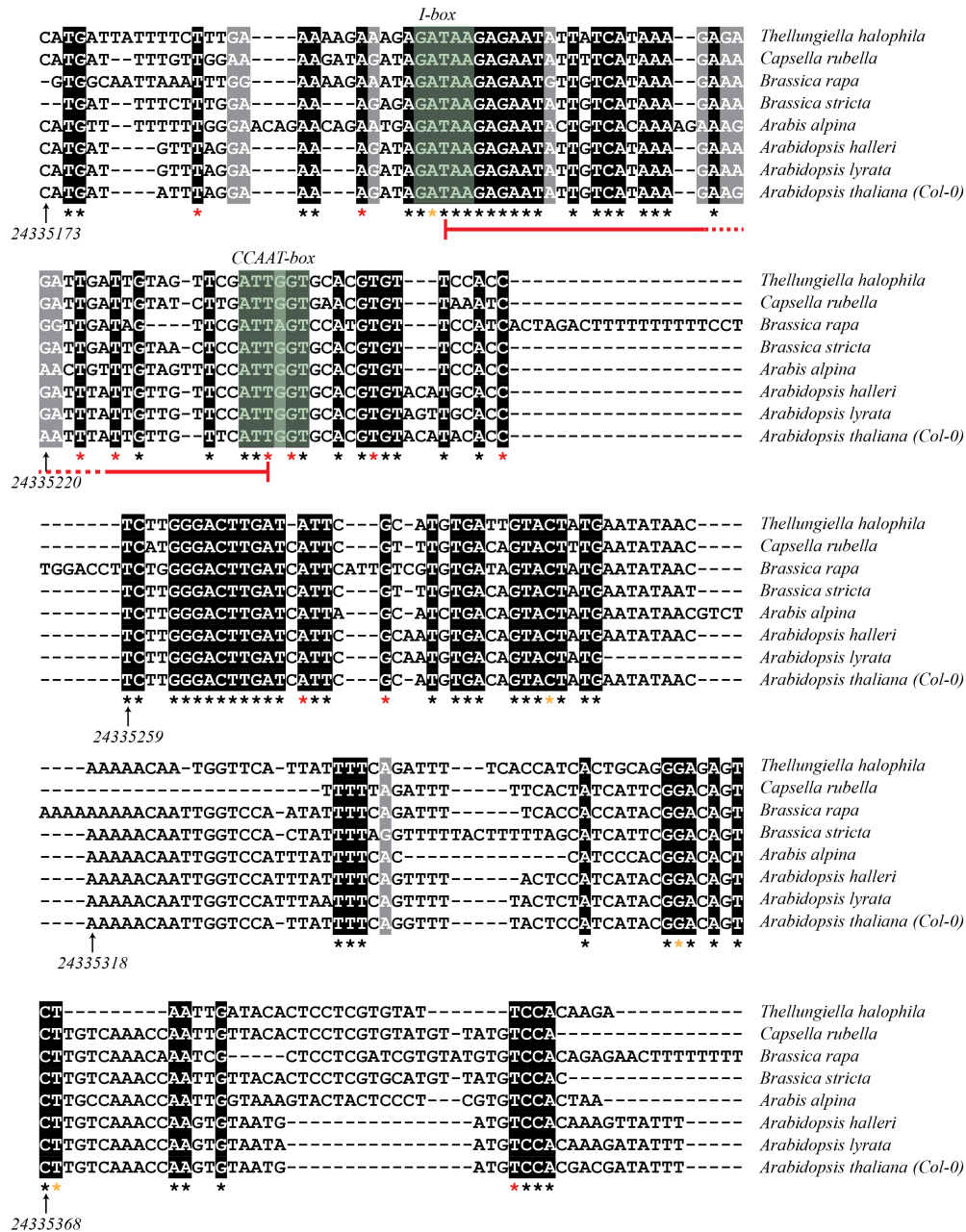


Figure 9: Natural variation in the novel enhancer (*Block E*) of *FT*.

Multiple alignments of *Block E* from different Brassicaceae species. The putative *CCAAT-box* and *I-box* TFBS are annotated in green. The multiple alignments were done with mVista and MUSCLE (Brudno et al. 2003; Edgar 2004; Frazer et al. 2004). The shadings were done with BOXSHADE, and correspond to the degree of conservation. Nucleotides conserved among species and the 1001 genomes project are annotated with a black asterisk. SNPs from the 1001 genomes project in nucleotides conserved among the Brassicaceae species tested are annotated with a red or yellow asterisk depending on the number of accessions sharing the SNP (< 10 and > 10 respectively). The red line corresponds to the deletion present in the accessions 6268 and 9442. The position on the Chromosome 1 is indicated by an arrow.

To investigate whether natural variation in the CREs nearby *FT* could alter *FT* expression and consequently the floral transition, I explored the 1001 genomes project for *FT* haplotypes with natural variation in the putative and functional TFBSs located at the photoperiodic control regions of *FT*.

After screening a global population of 1135 accessions, most of the candidates have a MAF below 0,5% and are therefore considered as “rare” (**Figure 10A**). The natural *CCAAT*-box variant of *FT*, previously described in the Ull2-5 accession, is also present in TV-10 (ID: 6258) and Ale-stenar-56-14 (ID: 997), two accessions coming from Sweden (Strange et al. 2011) (**Figure 10A**). The accessions Ömö2-1 (ID:7518) and Röd-17-319 (ID:1435), collected in Sweden, share a natural variant of the *E*-box located at *Block C*. Additionally, the accessions Lp2-2 (ID:7520) and Rovero-1 (ID:9976), originally from the Czech Republic and Italy, respectively, have two different SNPs in the *REalpha* element located at *Block A*. I also identified one natural deletion variant of *Block C* in the accession Rubezhone-1 (ID: 7323) and one of *Block E* present in the accessions TV-22 (ID: 6268) and Sim-1 (ID: 9442) (**Figure 10A**). The accession Rubezhone-1 shows the signature of a 65 bp deletion located upstream of the putative *I*-box at *Block C* (**Figure 8**). The accessions TV-22 and Sim-1 present the pattern of a 41 bp deletion spanning the putative *I*-box and *CCAAT*-box located at *Block E* (**Figure 9**).

To detect any significant associations between candidate *FT* haplotypes, containing non-coding sequence variation in the photoperiodic control region of *FT*, and geographical distributions, I applied a GLM model that take into account the population structure. The analysis does not reveal particular latitudinal and longitudinal clines (**Figure 10B**, **Supplementary Figure 4**).

To sum up, natural variation in highly conserved nucleotides at the photoperiod control regions of *FT* is rare in the global population of 1135 *Arabidopsis* accessions.

Results

A

Region	Element	Position	Variation	MAF	Accession
Block A	<i>E</i> -box	24331186	A>G	0,085	7518
		24331189	T>G	0,085	1435
	<i>Shadow3A</i>	24331238	G>C	0,427	6214/6216/6217/6218/9371
	<i>CORE2</i>	24331347	A>G	0,085	7223
Block C	?	2432600-24326065	Deletion	0,085	7323
	<i>Shadow2C</i>	24326143	A>G	0,684	4840/5023/6924 9314/9791/9806/9906/9911
	<i>Shadow3C</i>	24326163	G>A	0,034	7218/7337/7717/9739
	<i>Shadow4C</i>	24326179	G>A	0,085	5779
	<i>CCAAT</i> -box	24326211	G>A	0,256	997/6258/6974
	<i>REalpha</i>	24326244	G>A	0,085	7520
		24326245	G>C	0,085	9976
Block E	<i>CCAAT</i> -box	24335196-24335237	Deletion	0,171	6268/9442

B

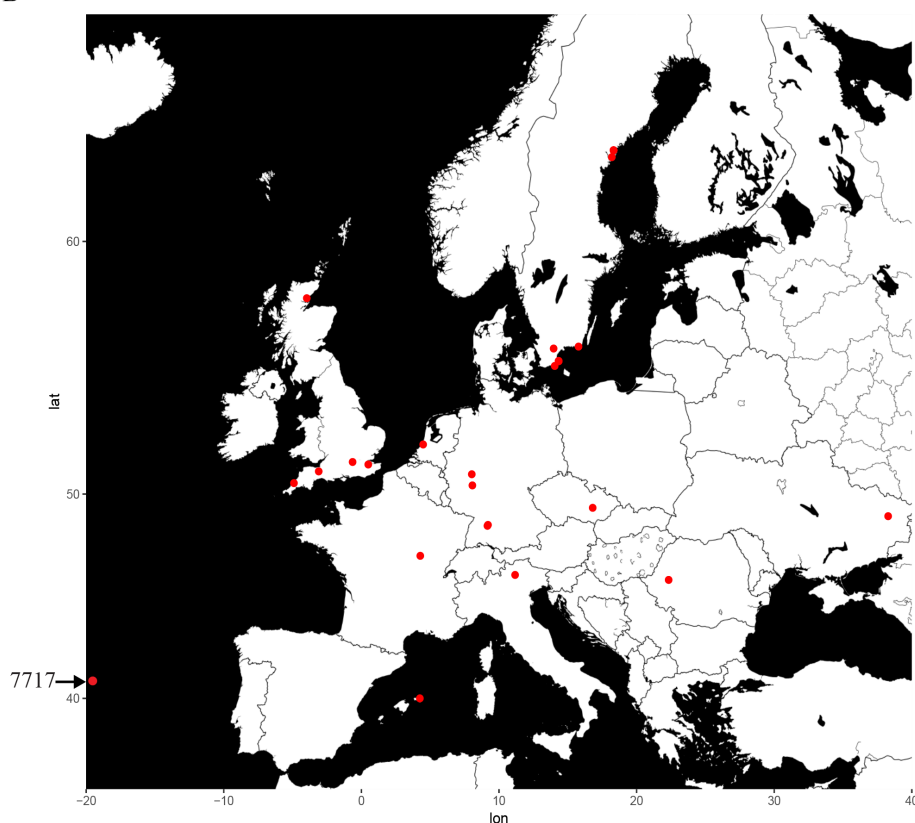


Figure 10: Natural variation in the photoperiod control regions of *FT* in *Arabidopsis*. (A, B) Summary table (A) and geographical distribution (B) of accessions with natural variation in the putative and functional TFBSs involved in the induction of *FT* expression. (A) MAF corresponds to the minor allele frequency in % in a global population of 1135 accessions provided by the 1001 genomes project. “?” indicates no putative TFBSs. (B) Red dots indicate the localization of the candidates. The accessions 7717 comes from the USA.

3.2.3 Transposable element polymorphisms at *FT*

Most of the eukaryotic genomes are composed of repetitive sequences resulting from the activity of TEs (Britten and Kohne 1968). The Arabidopsis genome, which is rather compact compared to the Maize genome, for example, has been shown to be composed of ~15% of TEs (de la Chaux et al. 2012). TEs are enriched in constitutive heterochromatin regions relatively devoid of genes compared to euchromatin regions (Copenhaver et al. 1999). TEs have the ability to move and to spread within a genome. Therefore active TEs are highly mutagenic, affecting protein function or gene expression for example. However, different epigenetic mechanisms evolved to repress TE expression and mobility (Slotkin and Martienssen 2007).

Transposable element polymorphisms (TEPs) are difficult to detect but present distinct signatures. Indeed, TE insertions induce split-read and discordant paired-read events that can be revealed using generic structural variation detection tools. (**Figure 11A**). I explored 1055 sequences from the 1001 genomes project revealing 16 accessions with TEP at the *FT* locus. (**Figure 11B**). Three accessions (T890 (ID: 6141), T780 (ID: 6131), and Sparta-1 (ID:6085)) coming from the South of Sweden present a TEP located upstream of *Block C*. These accessions share the same TEP, a ~2 kb VANDAL5/Mutator transposon (MuDr), suggesting an ancestral single insertion event. Seven accessions shows the signature of TEP located downstream of *Block C*. Among these seven accessions, Död2 (ID: 9352), Död3 (ID: 9353) and App1-16 (ID:5832), also coming from the South of Sweden, share a same TEP, a ~2 kb long MuDr, suggesting an ancient TE insertion event. The accessions OOE1-1 (ID: 15591) and OOE3-1 (ID: 15592), coming from Austria, present the pattern of a TEP downstream of *Block C*. Additionally, the accessions OOE1-1 (ID: 15591) and OOE3-1 (ID: 15592) are geographically close to each other suggesting an ancient TE insertion event. The accessions Dobra-1 (ID: 10018) and T580 (ID: 6115) present the signature of TEP. However, these accessions are geographically distinct from each other and seem to have different TE insertion events. The accessions Angel-1 (ID: 9980), Mammo-1 (ID: 9964) and Moran-1 (ID: 9967), coming from the South of Italy, share the same TEP, a ~ 250 bp SINE placed ~230 bp downstream of the 3'UTR of *FT*. The accessions Fjä2-6 (ID: 6022), Dju-1 (ID: 9343) and Sim-1 (ID: 9442) show the pattern of a TEP upstream of *Block E* (**Figures 11C and D**).

Results

The natural TEP variants of *FT* seem to be enriched in the South of Sweden. Therefore, I applied a generalized linear model (GLM) to detect any significant associations between *FT* haplotypes containing TEPs and geographical coordinates. After correction for population structure, the analysis does not reveal any particular latitudinal and longitudinal clines (**Figure 4D**).

In sum, using paired-end short reads data from the 1001 genomes project, four independent naturally TE insertions at *FT* were discovered. These non-annotated TEPs do not have the potential to alter the protein function or the mRNA splicing. However, their proximity to the photoperiod control and 3'UTR regions of *FT* raises the question about their effect on *FT* expression. Interestingly, all TEPs detected at *FT* are located in relatively open chromatin regions in the reference accession Col-0 (Zhang et al. 2012).

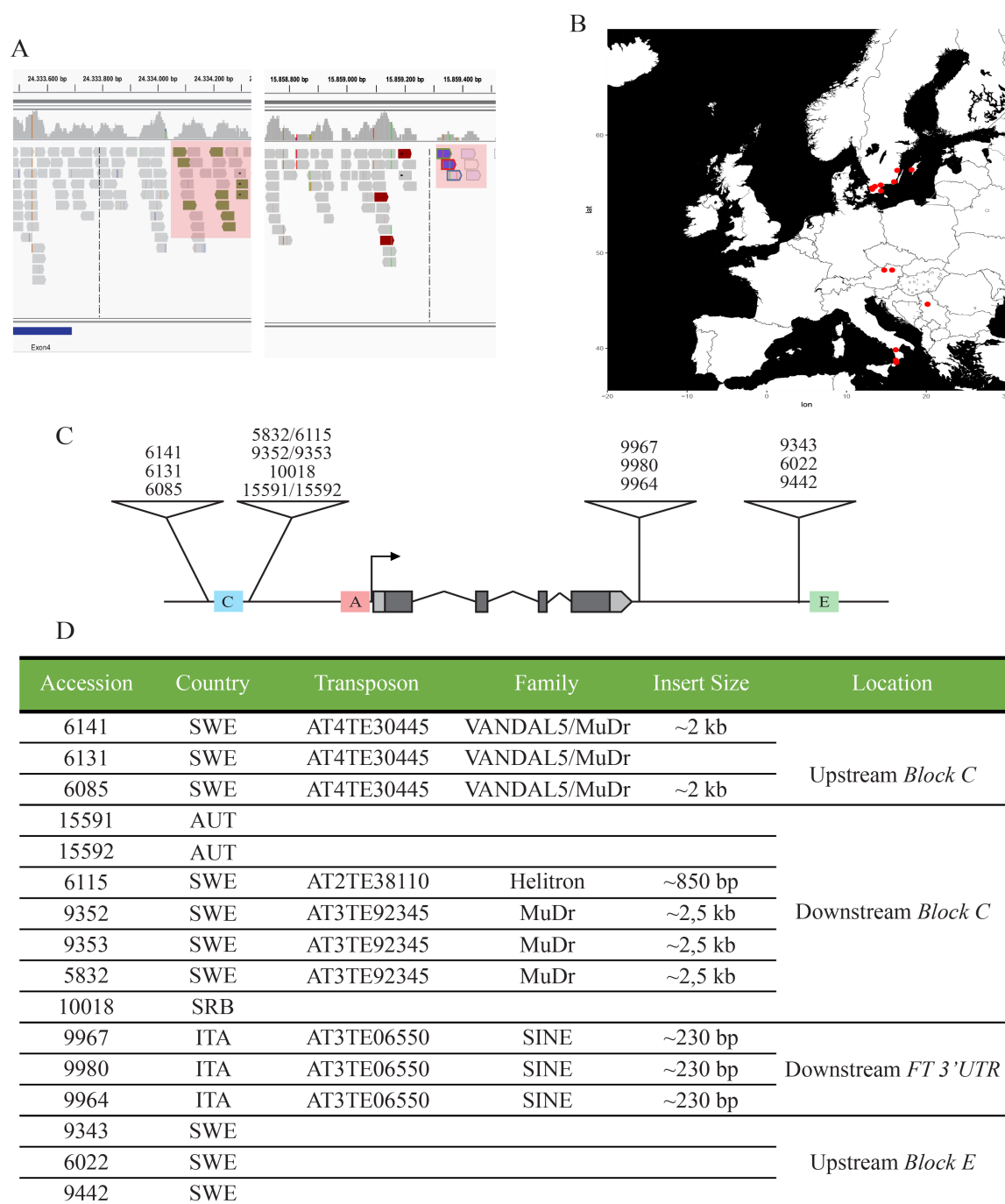


Figure 11: Screen for transposable element insertions at *FT* from paired-end short reads.

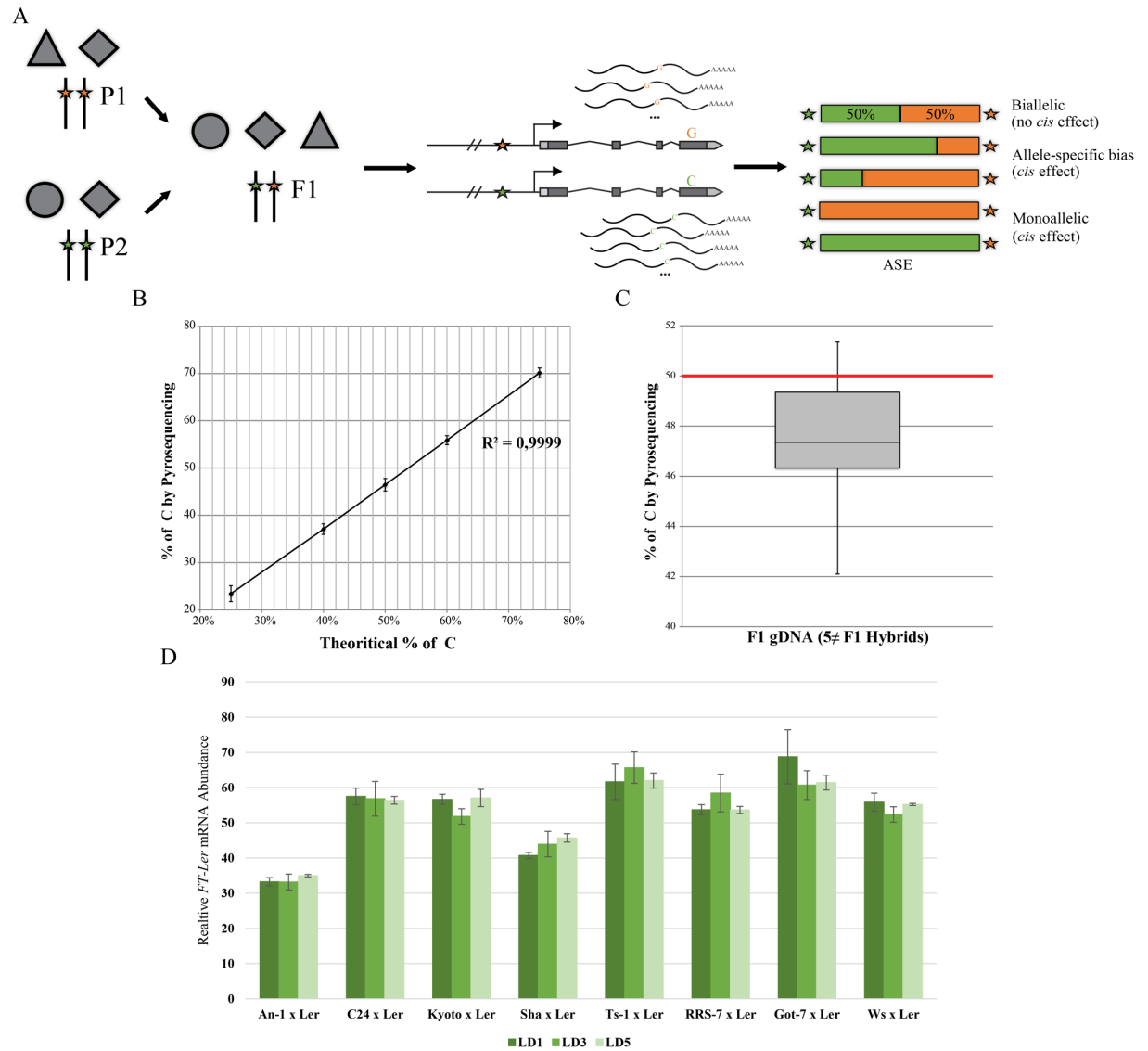
(A) Integrative genomics viewer (IGV) browser display of paired-end short reads aligned with BWA (Li and Durbin 2009; Robinson et al. 2011). Discordant paired-reads due to the TEP are highlighted in a red box. The reads mapping *FT* and the corresponding mates at the chromosome 3 are represented in green and blue, respectively. (B) Geographical distribution of the accessions containing a TEP at the *FT* locus. Red dots indicate the localization of the candidates. (C) Schematic representation of the TEP identified at the *FT* locus. The blue, red and green boxes correspond to *Blocks A, C* and *E*, respectively. Exons and UTRs of *FT* are highlighted in dark and light gray, respectively. The sense of the transcription is indicated by an arrow. (D) Table of accessions with a TEP at the *FT* locus. TE insertions at *FT* are checked by PCR when insert size indicated.

3.3 Characterization of the candidate non-coding variant *FT* haplotypes

3.3.1 Allele-specific expression of *FT* in F1 hybrids

Differences in a gene expression between accessions can be due to *cis* or *trans* effects. However, in heterozygotes, the parental alleles are exposed to the same *trans*-acting regulators, and therefore, differences in ASE reveal the effect of *cis*-regulatory genetic variation (Pastinen 2010) (**Figure 12A**). To estimate the accuracy of this assay, we quantified by pyrosequencing different gDNA ratios from two accessions (Col-0 and HR-5) selected for their SNP (C/G respectively) in the exon 4 of *FT*. Quantification of the ratios by pyrosequencing precisely reflects quantification of the ratios by quantitative PCR (qPCR), as the R-squared value of the trend line is 0.99992 (**Figure 12B**). However, we observed a bias in one of the quantification method since the abundance of C allele seems to be lower in the pyrosequencing assay compare to the qPCR quantification (**Figure 12B**). To evaluate this error without any pipetting variations, gDNA from the corresponding F1 hybrids was analyzed by pyrosequencing. The analysis confirmed an underestimation of the C allele by 2,75 % (**Figure 12C**). In subsequent pyrosequencing experiments, the error due to the underestimation of the C allele was corrected. In general, the pyrosequencing calibration experiment indicates a robust assay with negligible inter-PCR and sequencing variation.

The candidate and control accessions were crossed to the *Ler* accession based on the presence of a SNP in the exon 4 of *FT*. Because the parental accessions used in this assay have different vernalization requirements, the resulting F1 hybrids were first grown in SDs at 21°C for two weeks and then exposed to cold temperature in SDs for 12 weeks to repress *FLC* expression in every genetic background and therefore allow the detection of *FT* expression. After an additional two weeks in SDs at 21°C, the plants were then exposed to inductive photoperiod conditions (LD at 21°C). To set up a timeframe for ASE profile of *FT*, we collected the samples, one, three, and five days after the shift to photoperiod inductive conditions (LD1, LD3, and LD5, respectively). No significant differences were detected in the relative *FT-Ler* mRNA abundance in 9 different F1 hybrids (**Figure 12D**). Therefore, the further ASE profiles of *FT* were done at LD3 following the same experimental design.



Results

3.3.1.1 Allele-specific expression of *FT* in the rosette leaves

3.3.1.1.1 LD photoperiod

The analysis in F1 hybrids revealed four main groups of *FT* ASE profile (Figure 13A). The group “a” encompasses most of the F1 hybrids tested, where the candidate *FT* alleles were expressed at a relatively similar level compared to the reference *FT-Ler* allele. Every *FT* haplotype with TEP surrounding *Block C* (Död2, Död3, App1-16 and T580) fell in this category. These *FT* alleles are therefore as responsive as the *FT-Ler* allele to LD photoperiod induction. This trend was surprisingly also true for *FT-Est-1*, an *FT* haplotype previously associated by QTL mapping with differences in flowering time (Schwartz et al. 2009). In fact, the *FT-Est-1* mRNA transcripts levels were significantly less abundant than the *FT-Ler* allele compared to the gDNA control.

The group “b” embraces F1 hybrids with higher expression of the *FT* candidate allele compared to the *FT-Ler* allele and includes the An-1 (ID: 6898), Sorbo (ID: 6963), Sha (ID: 9805) and Kondara (ID: 6929) *FT* haplotypes. The *FT-An-1* allele shows the highest *FT*-mRNA level in this assay. Interestingly, a previous study detected a QTL at *FT* for the empty axil phenotype in multiparent RILs segregating the *FT-An-1* and *FT-C24* alleles. In this study, ASE analysis confirmed *FT-An-1* to be the more expressed allele while the presence of the weaker *FT-C24* allele enhanced the phenotype (Huang et al. 2013). Since the C24 accession contains the short *FT* promoter type whereas the Ler-0 and An-1 accessions have the medium *FT* promoter type, it had been concluded that the *FT-C24* allele was less responsive to photoperiod. However, my experiment indicates that the *FT-An-1* allele is more responsive to photoperiod independent of the *FT* promoter type.

At the opposite, the group “c” encompasses F1 hybrids containing *FT* candidate alleles with lower *FT* expression compared to the *FT-Ler* allele. This group includes *FT* haplotypes selected for SNPs in highly conserved elements located in *Blocks A* and *C* (*REalpha*, *Shadow3A*, and *E-box*). It reveals a weaker photoperiodic response in the accessions TFÄ08 (ID: 6218), TFÄ07 (ID: 6217), FäL1 (ID: 9371), Litva (ID: 7236), Lp2-2 (ID: 7520) and Vinslöv (ID: 9057).

Finally, the groups “d”, “e” and “f”, encompass F1 hybrids with a mono-allelic expression of the *FT-Ler* allele. These groups include the *FT-Ull2-5*, *FT-Ale-stenar-56-14* and *FT-TV-10* alleles, three candidate *FT* haplotypes with a SNP in the functional *CCAAT*-box located at *Block C*. The analysis of ASE of *FT* reveals that the three accessions are strongly affected in the photoperiodic response and follow previous observations done in the Ull2-5 accession (Strange et al. 2011). Indeed, as mentioned before, the *FT-Ull2-5* haplotype has been reported to affect flowering time in F2 populations. As a full deletion of the functional *CCAAT*-box has been shown to have a radical impact on *FT* expression, it would be easy to speculate that the SNP shared by the Ull2-5, Ale-stenar-56-14 and TV-10 accessions, is the causal natural variation. However, the presence of additive natural variations in this *FT* haplotype cannot be excluded.

3.3.1.1.2 Ambient temperature

A previous study reported induction of *FT* expression at dusk under SDs at 27°C (Balasubramanian et al. 2006). This induction requires PIF4, PIF5 as well as CO proteins (Fernández et al. 2016). Because CO especially boosts *FT* expression in a *Blocks A* and *C* dependent manner, the *FT* haplotypes characterized could react in a similar fashion to LD photoperiod conditions. Analysis of ASE of *FT* in the rosette leaves of F1 hybrids confirmed a trend comparable to photoperiod inductive conditions in my assay for most of the tested F1 hybrids (**Figure 13B**). Interestingly, the *FT* haplotype carrying the natural *CCAAT*-box variant, shared by the *Ull2-5*, *TV-10* and *Ale-stenar-56-14* accessions, remain strongly affected. Note, however, that the previous group “b” is closer to a bi- rather than a mono-allelic expression of *FT* in these conditions.

Results

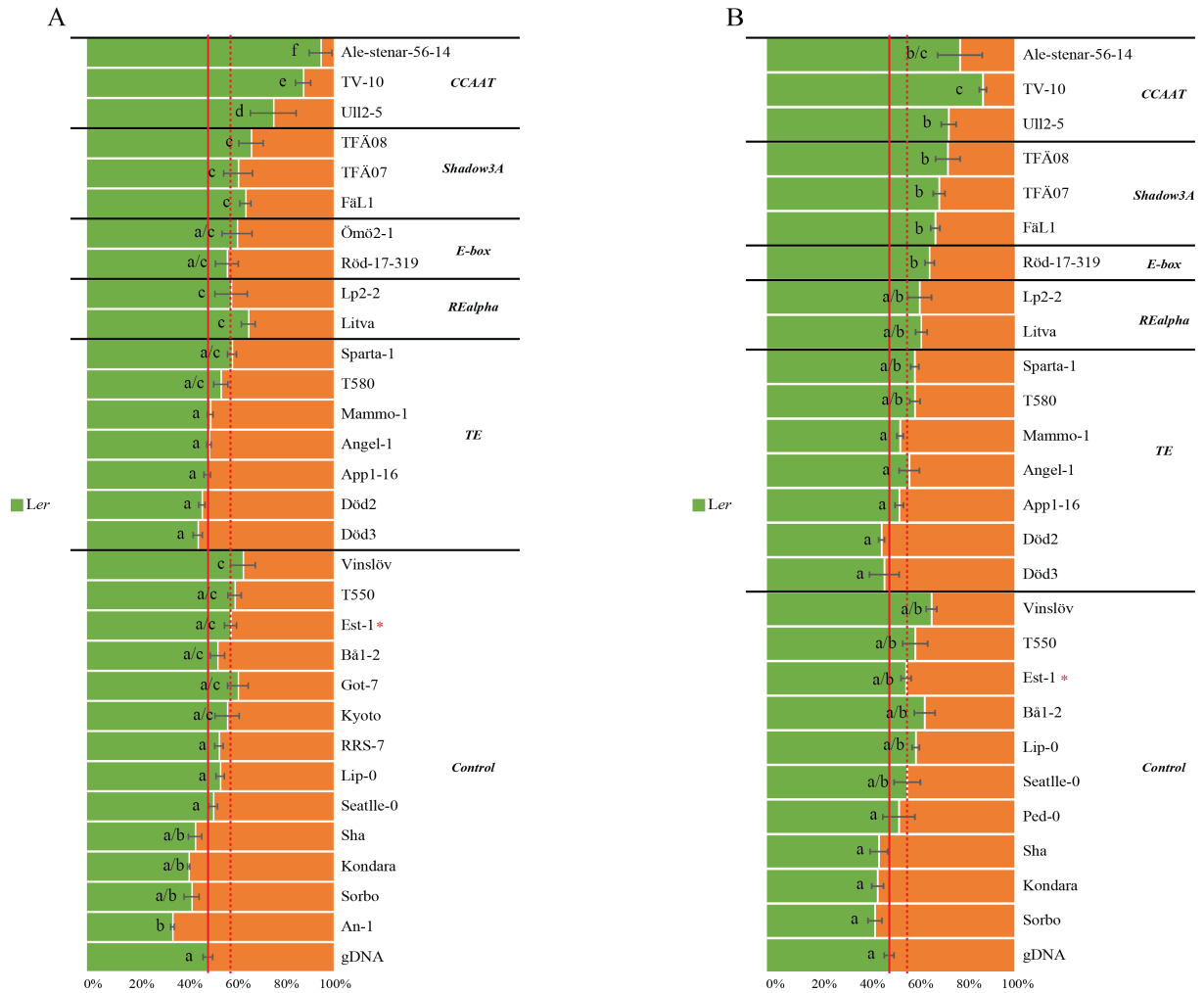


Figure 13: ASE of *FT* in the rosette leaves in response to photoperiod and ambient temperature.

(A, B) ASE of *FT* in the rosette leaves of F1 hybrids induced in photoperiod inductive condition (LD at 21°C) (A) and ambient temperature inductive condition (SD at 27°C) (B). *Ler* was used as a recurrent mother. The relative amount of *Ler* and candidate *FT* mRNA are represented in green and orange, respectively. The ASE is denoted as the mean \pm SE ($n \geq 6$). Statistical significance was determined using a one-way ANOVA followed by a Holm-Sidak correction for multiple comparisons. Significant differences are indicated by different letters ($P < 0.05$). Quantification of parental *FT* alleles in F1 hybrids gDNA is shown as control and represented with a red bar. * represents F1 hybrids containing the *FT-Est-1* haplotype previously associated with differences in flowering time (Schwartz et al. 2009). The relative amount of *FT-Est-1* mRNA is annotated with a red dashed bar.

3.3.1.2 Allele-specific expression of *FT* in the siliques

Most of the *FT* haplotypes tested in our assay were selected for the presence of natural variation in putative and functional TFBSs located at the photoperiod control regions of *FT*. Our group demonstrated *FT* expression to be independent of *Block C* in the siliques (Liu et al. 2014b). However, preliminary results suggest an induction of *FT* expression in the siliques at zeitgeber time (ZT) 16 in LDs (Fabian Bratzel, unpublished data). This observation led us to investigate how the selected *FT* haplotypes would react in developing siliques of F1 hybrids. Unexpectedly, all F1 hybrids tested showed a balanced bi-allelic *FT* expression in photoperiod inductive conditions (**Figure 14**). This trend was also true for *FT-Ull2-5* and *FT-Ale-stenar-56-14*, two *FT* haplotypes dramatically affected in the photoperiodic induction of *FT* expression in rosette leaves. Taken together, the data indicate that the less responsive *FT* haplotypes in the rosette leaves remain as sensitive to LDs as the *FT-Ler* control haplotype in the siliques.

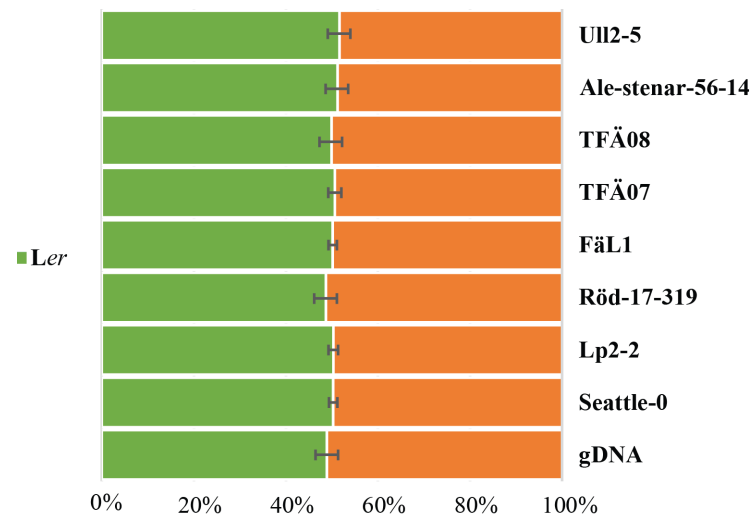


Figure 14: ASE of *FT* in the siliques in response to photoperiod.

(A, B) ASE of *FT* in the developing siliques of F1 hybrids induced in photoperiod inductive condition (LD at 21°C). *Ler* was used as a recurrent mother. The relative amount of *Ler* and candidate *FT* mRNA are represented in green and orange, respectively. The ASE is denoted as the mean \pm SE ($n \geq 6$). Statistical significance was determined using a one-way ANOVA followed by a Holm-Sidak correction for multiple comparisons. No significant differences were detected ($P < 0.05$).

3.3.2 Complementation assay

To characterize the effect on the floral transition of the less responsive *FT* haplotypes characterized in the previous ASE assays, I analyzed the ability of different constructs containing the candidate *FT* promoters to complement the *ft-10* mutant phenotype. The genomic region of *FT*, beginning from the start codon to 2.3 kb downstream of the stop codon, was fused to the *FT* promoter variants of Ale-stenar-56-14, TFÄ08, FäL1 and Col-0. The 2.3 kb genomic region downstream of *FT* contains the novel enhancer *Block E* (Johan Zicola, Thesis). Additionally, I also introduce point mutations into the *Ale-stenar-56-14p::FTgDNA* and the *TFÄ08p::FTgDNA* constructs to evaluate whether the SNPs present in the *CCAAT-box* and the *Shadow3A cis-elements* are the causal natural variations for the asymmetric ASE of *FT* observed previously.

Surprisingly, the floral transition of the control *Col-0p::FTgDNA ft-10* plants was highly variable between the T3 independent lines (**Figure 15**). While the lines #1 and #2 have a total leaf number before flowering comparable to wild type plants, the flowering time of the lines #4 and #5 is similar to *ft-10* mutants. Furthermore, this trend is also true for *ft-10* lines transformed with the *TFÄ08p::FTgDNA*, *TFÄ08(funct)p::FTgDNA* and the *Ale-stenar-56-14(funct)p::FTgDNA* constructs (**Figure 15**). Nevertheless, the *Ale-stenar-56-14p::FTgDNA* and *FäL1p::FTgDNA* constructs could not complement the *ft-10* mutant phenotype in the lines tested in my assay, suggesting that the *FT* promoter variants, present in the Ale-stenar-56-14 and FäL1 accessions, contain natural variations that affect the floral transition at least in the Col-0 background (**Figure 15**). However, the high variation in flowering time of *ft-10* plants transformed with the different constructs generated brings to question the reliability of my assay.

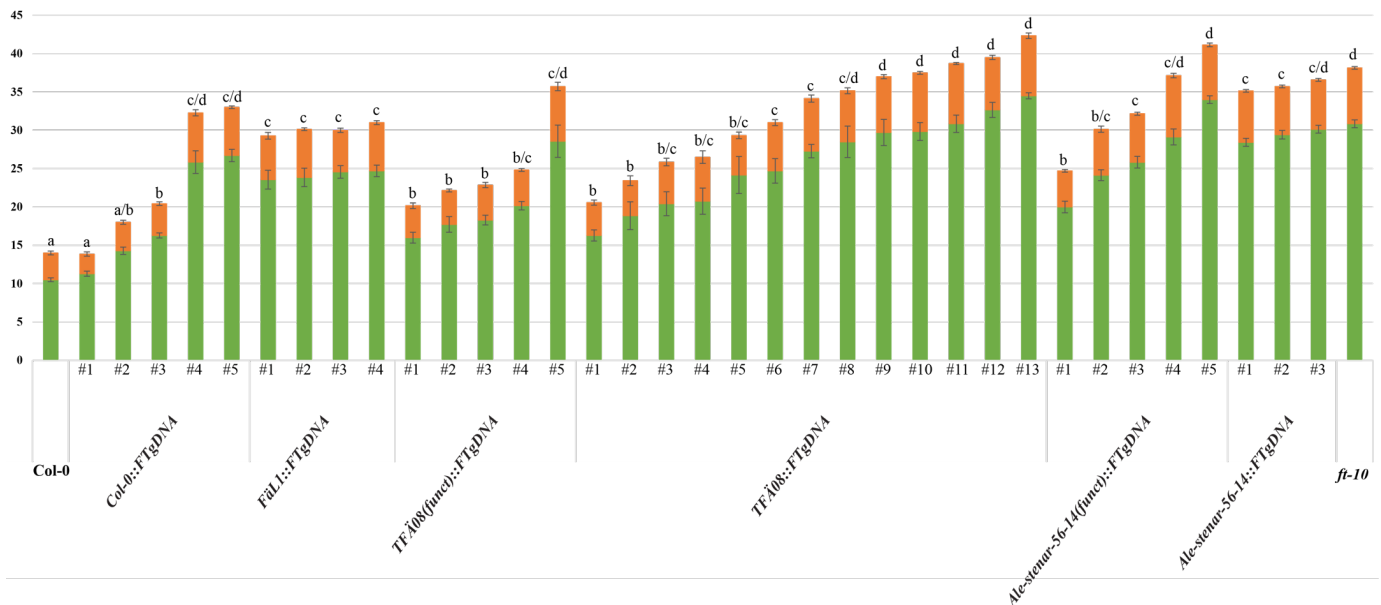


Figure 15: *FT* promoter variants in response to LDs in a Col-0 genetic background. Flowering time of *ft-10* mutants in T3 generation carrying transgenic constructs driving *FT-gDNA* by a 5,7 kb long *FT* promoter from the different candidate accessions. Plants were grown in LDs. Col-0 and *ft-10* plants were used as controls. Number of rosette and cauline leaves, respectively in green and brown, are shown as the mean \pm SE ($n = 10$). Statistical significance was determined using a one-way ANOVA followed by a Holm-Sidak correction for multiple comparisons. Significant differences are indicated by different letters ($P < 0.05$).

The position of the transgene could explain the high variation of flowering time observed in T3 transgenic lines (Shi et al. 2010). Indeed, the chromatin context of the T-DNA insertion site could, for example, influence gene expression. However, several studies reported the insertion of tandem-repeated T-DNAs (Gelvin 2003). For that purpose, I selected randomly three to four independent T3 transgenic lines for each construct and estimated the transgene copy number based on a qPCR approach.

Surprisingly, the qPCR results indicate the presence of T3 lines with a high transgene copy number. The identification of transgene copy number seems to be a robust assay for low copy number since the qPCR results reflected the transgene copy number of three independent transgenic lines previously genotyped by Southern Blotting as a single copy of the transgene (**Figure 16A**). However, the estimation of the transgene copy number seems to be coarse for transgenic lines with more than three copy as the standard error indicates (**Figure 16A**). The data suggest a strong bias with the transgene copy number variation since for each construct, the earliest flowering *ft-10* plants transformed correspond to lines with high copy number of the transgene (**Figure 16B**).

Results

Considering the “ α ” group, which corresponds to transgenic lines with a single copy of the transgene, we noticed that the *ft-10* transgenic lines #3 and #4 transformed with the *Col-0-5,7kbFTp::FTgDNA* construct, have significantly different flowering behavior in LDs (**Figure 16B**). This observation could be explained by the position of the transgene. However, the global population of lines carrying the *Ale-stenar-56-14p:FTgDNA* and *FäL1p::FTgDNA* transgene flowered significantly later than with the *Col-0-5,7kbFTp::FTgDNA* transgene control, which could indicate the presence of causal natural variation in the *FT* promoter of the *Ale-stenar-56-14* and *FäL1* accessions (**Figure 16B**). A higher number of lines with a single copy of the transgene is required to evaluate the position effect of the transgene.

Taken together, the data reveal the presence of transgenic lines with high transgene copy number in a T3 population despite a segregation following a ratio of 3:1 in the previous T2 generation. Therefore, this observation indicates the presence of tandem-repeated T-DNAs inserted at a single locus. After screening for copy number variation of the transgene, the data suggest that the *Ale-stenar-56-14p::FTgDNA* and *FäL1p::FTgDNA* constructs are impaired in their ability to complement the *ft-10* mutant phenotype compared to the *Col-0-5,7kbFTp::FTgDNA* control. However, we obviously need a bigger population of T3 transgenic lines that carry only one copy of the transgene to exclude the effect of transgene copy number variation and estimate the position effect.

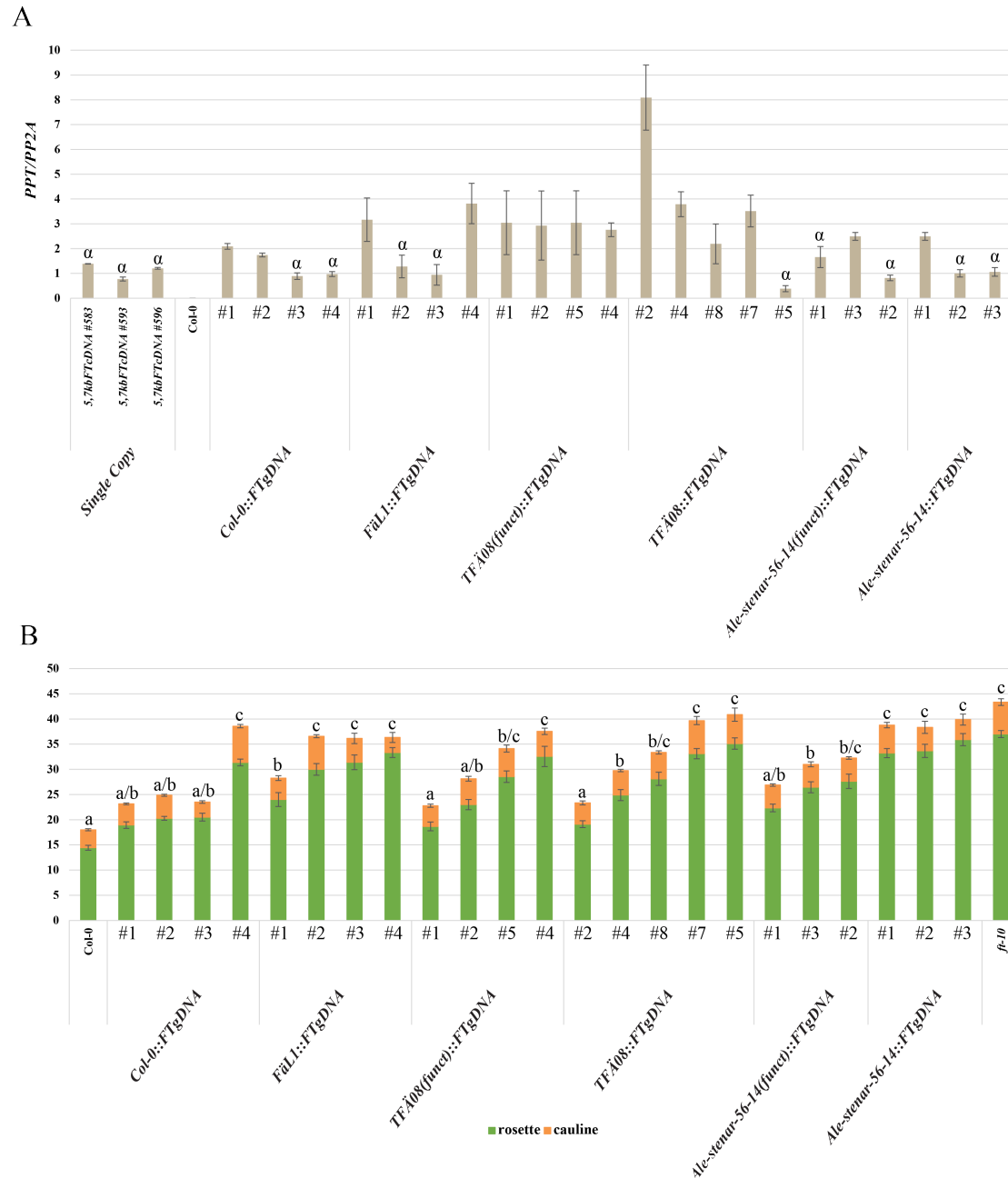


Figure 16: Screening for tandem repeat insertions in complementation lines.

(A) qPCR analysis on gDNA of *PP2A* and *PPT* from *ft-10* seedlings carrying the different natural *FT* promoter variants constructs. Values represent mean \pm SE of *PPT* relative to *PP2A* levels for three technical replicates. Single copy lines were genotyped by Southern Blotting and used as controls. Col-0 seedlings were used as a negative control. (B) Flowering time of *ft-10* mutants in T3 generation carrying transgenic constructs driving *FT-gDNA* by a 5,7 kb long *FT* promoter from the different candidate accessions. Plants were grown in LDs. Col-0 and *ft-10* plants were used as controls. Number of rosette and cauline leaves, respectively in green and brown, are shown as the mean \pm SE ($n = 10$). (A, B) Statistical significance was determined using a one-way ANOVA followed by a Holm-Sidak correction for multiple comparisons. Significant differences are indicated with α and different letters in A and B, respectively ($P < 0.05$).

3.4 Flowering behavior of the parental accessions

Since the effect of the natural *CCAAT*-box *FT* variant is the strongest in ASE of F1 hybrids, we had a particular interest in the Ale-stenar-56-14, TV-10 and Ull2-5 accessions. These three accessions were geographically close to each other in the South of Sweden (**Figure 17A**). Interestingly, they belong to the latest flowering plants in a GWAS study carried out in a global population of *Arabidopsis* accessions exposed to 10°C and 16°C (**Figures 17B and C**). The Ull2-5 accession has been reported to have a strong *FLC* haplotype causing a particularly slow vernalization response (Li et al. 2014). Previous field experiments done in the North of Sweden (close to Lövvik) nicely demonstrated a complete vernalization response after 12 weeks of growth preceding winter. This trend was true for the major *FLC* haplotypes that require an extensive period of cold for a full vernalization response (Duncan et al. 2015). Surprisingly, overwintered plants in this field site bolted within two weeks upon snowmelt including the Ull2-5 accession. In addition to field experiments, the floral transition of mutants impaired in the photoperiodic response has been reported to decrease with exposure to cold temperature (Martinez-Zapater and Somerville 1990; Koornneef et al. 1991; Ausin et al. 2004). Taken together, these observations led us to compare the flowering behavior of the candidate accessions exposed to 12 weeks of cold treatment in SDs with *ft* null mutants. Two different null allelic mutants of *FT*, *ft-10* and *ft-CRISPR*, as well as *Col-FRI_{sf2}* plants were also included as controls. Consistent with previous reports (Duncan et al. 2015), the flowering time of the tested accessions was similar to *Col-FRI_{sf2}* plants, including the Ale-stenar-56-14 and TV-10 accessions. However, unexpectedly, the two different null allelic mutants of *FT* flowered with the same number of leaves compared to wild-type plants under these conditions (**Figure 17D**). The data revealed that the effect of a null allele of *FT* could be bypassed after extensive exposure to cold treatment.

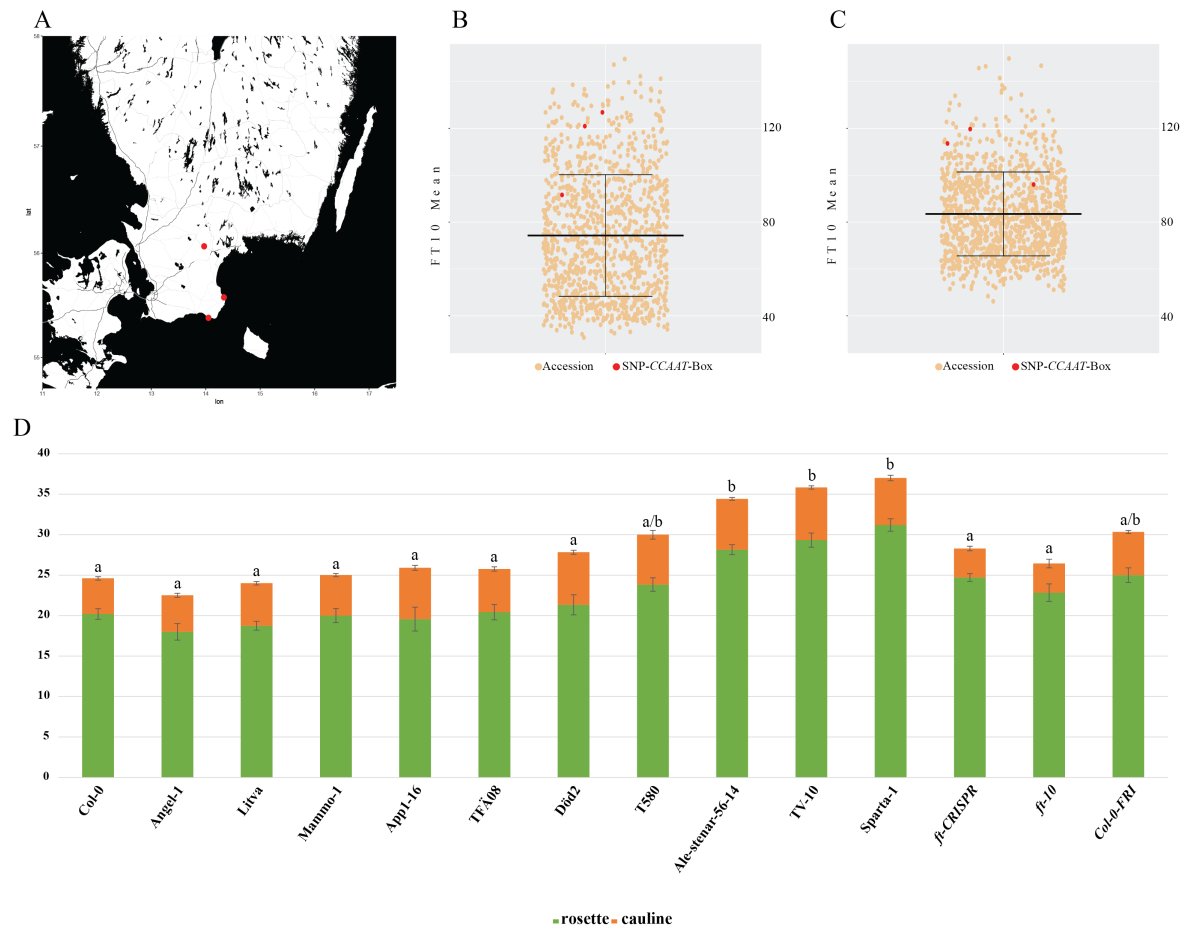


Figure 17: Flowering time of the candidate accessions after 12 weeks of cold temperature exposure.

(A) Geographical distribution of the natural *CCAAT*-box variant accessions Ull2-5, Ale-stenar-56-14 and TV-10. (B, C) Flowering time of 1163 accessions scored in days until the first flower open with the mean \pm SD. Plants were exposed to LDs at 10°C (B) and 16°C (C) (The 1001 Genomes Consortium 2016). The natural *CCAAT*-box variant accessions Ull2-5, Ale-stenar-56-14 and TV-10 are depicted in red. (D) Flowering time of the accessions selected with the *ft-CRISPR*, *ft-10*, and *Col-Sf2* as controls. The plants were successively grown for two weeks in SDs, 12 weeks of cold treatment in SDs and finally transferred in LDs. Rosette and cauline leaves respectively in green and orange represent the mean \pm SE ($n \geq 8$). The statistical significance was determined using a one-way ANOVA followed by a Holm-Sidak correction for multiple comparisons. Significant differences are indicated by different letters ($P < 0.05$).

3.5 Characterization of the impact of *FT* after extensive cold exposure

Since the *FT* protein has been shown to be the main component of the florigen signal at least in the Col-0 background, the total leaf number of *ft-10* mutants grown for 12 weeks in the cold compared to wild type plants raised the question of whether the florigen remains crucial in this floral transition. To investigate this question further, *ft-10* and *tsf-1* single mutants were grown for 6 and 12 weeks of cold temperature in SDs together with *ft-10 tsf-1* double mutants and wild type plants. 6 and 12 weeks of cold exposure strongly reduced the effect of the *ft-10* mutant allele on the total leaf number with respectively ~ 10 and ~ 3 leaves more than wild-type plants (**Figures 18A and C**). Interestingly, 6 and 12 weeks of cold temperature exposure also shorten the bolting time of *ft-10* mutants but with a milder effect compared to the decrease in total leaf number (**Figure 18B and C**). *tsf-1* mutants did not show a highly significant delay in flowering time compared to wild-type plants under all conditions (**Figure 18A, B and C**). However, *ft-10 tsf-1* double mutants showed a significantly higher number of both total leaves and days to bolting compared to *ft-10* mutants revealing the impact of *TSF* on the floral transition under these experimental conditions (**Figure 18A, B and C**). Additionally, the total leaf number of *ft-10 tsf-1* double mutants significantly decreased after 12 weeks of vernalization (**Figure 18B**). Taken together, the data indicate that the florigen has a weaker impact on the floral transition after extensive exposure to cold temperature.

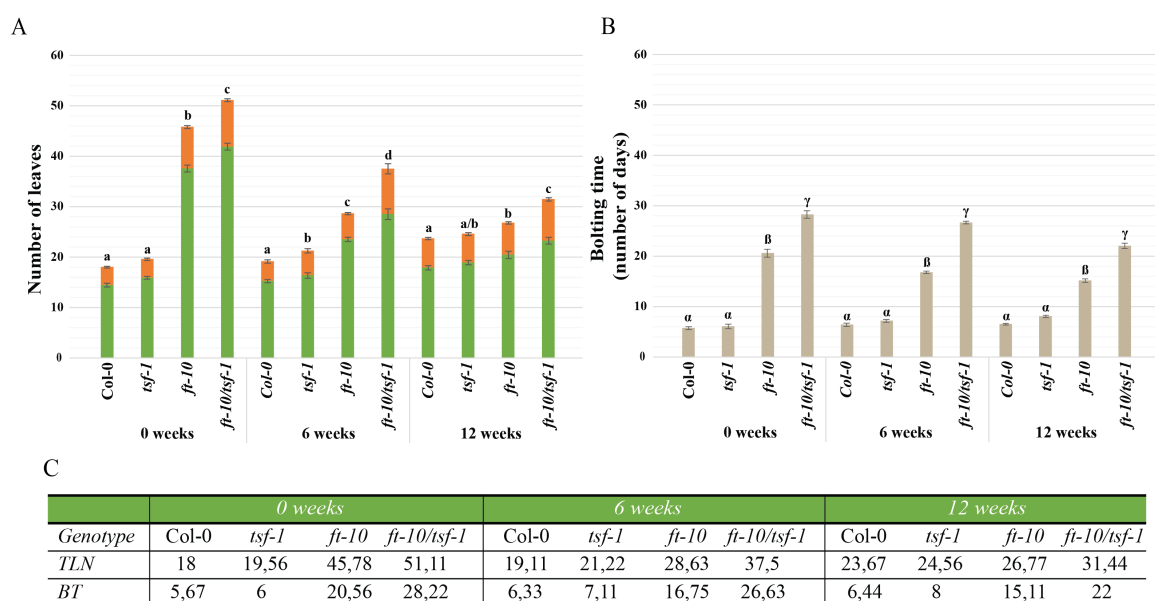


Figure 18: Characterization of the impact of the florigen on flowering time after extensive vernalization.

See following page.

Figure 18: Characterization of the impact of the florigen on flowering time after extensive vernalization.

(A, B) Flowering time of Col-0, *tsf-1*, *ft-10*, *ft-10 tsf-1* in LDs after different periods of cold temperature exposure. The plants were successively grown for 2 weeks in SDs, for different periods (0, 6 and 12 weeks) of cold treatment in SDs and finally transferred in LDs. Rosette and cauline leaves respectively in green and orange in (A), bolting time after transfer in LDs in brown represent the mean \pm SE ($n \geq 8$). For each cold treatment, the statistical significance was determined using a one-way ANOVA followed by a Holm-Sidak correction for multiple comparisons. Significant differences are indicated by different letters ($P < 0.05$). (C) Summary results of the flowering time. TLF and BT corresponds to the average of the total leaf number (rosette + cauline leaves) and the bolting time after transfer to LDs, respectively.

3.6 Characterization of winter Arabidopsis accessions

To investigate whether slow vernalization behavior accessions would carry more functionally differentiated *FT* haplotypes, I characterized *FRI* in a global population provided by the 1001 genomes project as a first screen. Indeed, *FRI* has been shown to play a central role in the vernalization response. Single season accessions are often homozygous for a nonfunctional *FRI* allele. Most of the naturally occurring variations at *FRI* that affect flowering time in Arabidopsis are due to structural variations. The reference accession Col-0 is a good example as it contains a nonfunctional *FRI* allele due to a 16 bp deletion present in the first exon of *FRI*. This deletion leads to a frameshift and the apparition of a premature stop-codon that ultimately affect the function of the protein. As Col-0 is used as a reference in genome mappings, the characterization of *FRI* remains challenging. Indeed, the corresponding deletion is discarded during the reads mapping and is not annotated in the 1001 genomes project. The analysis of short paired-end sequences remains the most straightforward method to detect structural variations at *FRI*. The *FRI* allele of *Ler* has been shown to have a weaker effect on the floral transition compared to the fully functional *FRI-Sf2* allele due to a reduction in *FRI* expression. Indeed, this allele has a deletion of 376 bp in the promoter of *FRI* that span the TSS (**Figure 19A**). Compared to the *FRI-Sf2* allele, the *FRI-Ler* allele has been shown to induce lower *FLC* expression leading to reduced vernalization requirement (Schmalenbach et al. 2014). Additional *FRI* haplotypes containing a premature stop-codons have been shown to encode nonfunctional *FRI* proteins (Shindo et al. 2005).

After filtering out all accessions with a sequence coverage lower than 6 reads, I finally analyzed a total of 995 accessions. In this population, I identified 73% of *FRI-Sf2* haplotype

Results

like, that are more likely to be functional. In the remaining nonfunctional *FRI* haplotypes, the *FRI-Ler* and the *FRI-Col-0* haplotypes represent respectively 15% and 11% of the population studied. The haplotypes containing a premature stop-codon represent only 1% of the population analyzed in our study (**Figure 19B**). The geographical distribution of the characterized *FRI* haplotypes suggests an enrichment of putative functional *FRI* haplotypes in Sweden and, surprisingly, in Spain. Most of the nonfunctional *FRI* haplotypes seem to be located in Central Europe (**Figure 19C**). To sum up, the high number of putative functional *FRI* alleles indicate that the slow vernalization behavior is more likely not to be explained by *FRI*. It would be, therefore, interesting the natural variation at *FLC*.

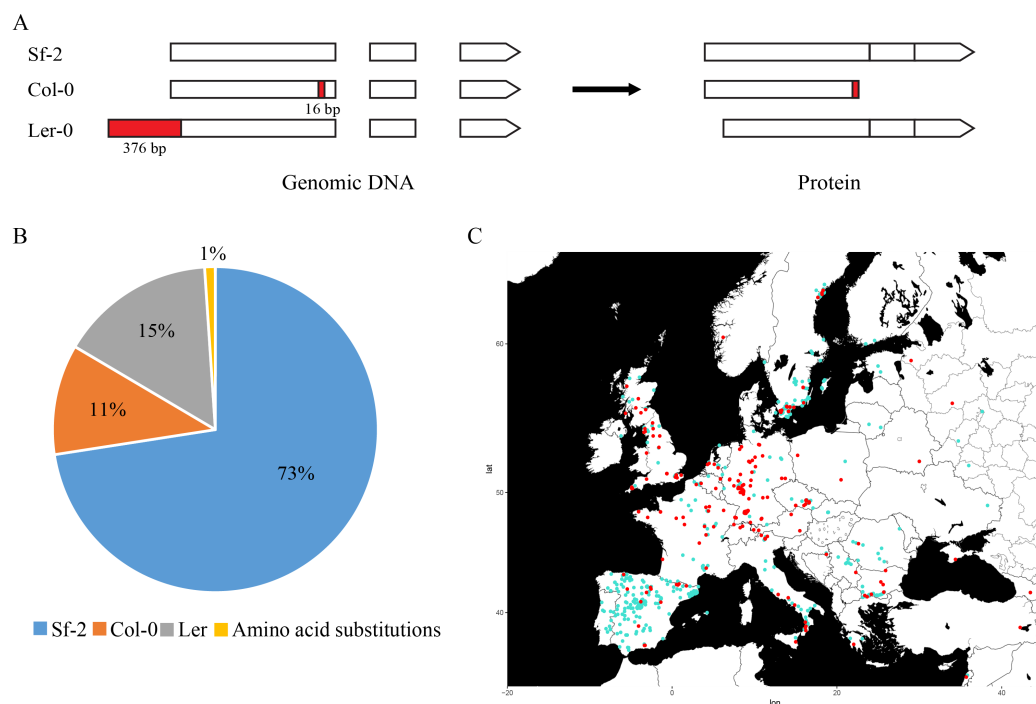


Figure 19: Characterization of *FRI* in the 1001 genomes project.

(A) Schematic representation of the main structural variations at *FRI*. The genomic sequences and the corresponding predicted proteins are represented on the left and the right respectively. Regions in red indicate the deletions. The *FRI-Sf-2* allele is functional. The *FRI-Col-0* allele is nonfunctional due to a deletion of 16 bp in the first exon leading to a frame shift and a premature stop-codon. The *FRI-Ler* allele is weak due to a deletion of 376 bp in the proximal promoter of *FRI* spanning the first exon. (B) Characterization of *FRI* in the 1001 genomes project with Illumina paired-end read sequences. Every accessions with a sequence coverage lower than 6 reads were filtered out. Accessions showing exclusively the structural variation of 16 bp are categorized in the *FRI-Sf-2* haplotype. Accessions with no structural variation of 16 bp are classified in the *FRI-Col-0* haplotype. Accessions with the structural variation of 376 bp are considered as the *FRI-Ler* haplotype. Accessions with nonsense mutations are classified in the amino acid substitution group. (C) Geographical distribution of the accessions characterized for *FRI*. Accessions with the *FRI-Sf2* haplotype are considered as functional and depicted in turquoise. The group of nonfunctional *FRI* haplotypes are shown in red.

4. Discussion

4.1 Natural variation at *FT* in *Arabidopsis*

Naturally-occurring genetic variation at *FT* has been uncovered by GWAS and QTL mapping with differences in flowering time (Schwartz et al. 2009; Li et al. 2010; Strange et al. 2011; Brachi et al. 2013; Sanchez-Bermejo and Balasubramanian 2015; The 1001 Genomes Consortium 2016). Although the causal natural variations remain unknown, the flowering time QTLs closely linked to *FT* are exclusively mapped to non-coding regions.

Consistent with these observations, the analysis of sequences of a global population composed of 1135 *Arabidopsis* accessions provided by the 1001 genomes project, revealed that naturally-occurring genetic variations that could affect FT protein function or structure are uncommon. While no structural variation was found in the coding sequence of *FT*, nine nsSNPs may change the protein sequence of FT. These nsSNP *FT* haplotypes have a MAF below 0,5% and are therefore considered as rare. Based on amino acid substitution predictions, only four nsSNPs affect amino acids highly conserved among 88 FT homologs of flowering plants and are therefore likely to modify the protein function or structure of *FT* (Ho and Weigel 2014). Additionally, a nsSNP alter the stop-codon of *FT* and leads potentially to a longer FT protein (**Figure 4**). Taken together, despite the fact that the functionality of these nsSNPs has to be confirmed, the strongest candidate *FT* haplotypes with possible altered FT protein structure or function are unique in a global population of 1135 accessions. The pleiotropic effect of *FT* could explain the high conservation of its coding sequence. Indeed, *FT* has been shown to be involved in a wide range of developmental processes in *Arabidopsis*, including flowering time, floral reversion, fruit set and stomatal opening (Kinoshita et al. 2011; Chen et al. 2014; Liu et al. 2014b).

Non-coding genetic variation at *FT* is more likely to explain the peaks of association identified at *FT* in the different flowering time GWAS. Indeed, while the essential function of the FT protein remains intact, non-coding variations would affect the level of *FT* mRNA in specific developmental processes and therefore potentially contribute to adaptation. Despite the fact that the characterization of active TFBSs is challenging, the combination of computational and evolutionary approaches identified *FT* haplotypes with natural

variations in putative and functional TFBSs located at the photoperiodic control regions of *FT*. Most of the non-coding *FT* haplotypes identified in this screen have a MAF below 0,5% in a global population of 1135 accessions (**Figure 10**). Additionally, the global distribution of the non-coding *FT* haplotypes does not seem to reveal any latitudinal or longitudinal clines.

This observation suggests that the functional units of the CREs involved in the photoperiodic induction of *FT* expression in the rosette leaves are globally preserved from natural variation. However, further characterization of active TFBSs involved in the photoperiodic induction of *FT* expression is required to detect any differences in the selective processes that occurs at *FT*.

The functional structural variations affecting the promoter length of *FT* are the most common *FT* haplotypes in *Arabidopsis* (Liu et al. 2014a). These *FT* haplotypes, previously associated with differences in seed production and ultimately fitness, have been associated with latitudinal clines. However, the analysis of paired-end short reads revealed a geographical distribution that contrast the previous latitudinal trends previously observed. Indeed, the medium and long *FT* promoter types seem to accumulate with Northern and Southern latitudes, respectively. Additionally, the short *FT* promoter type, previously associated with Southern latitudes is also present in Sweden (**Figure 5**). Therefore, it would be interesting to further investigate the ecological significance of the *FT* promoter variants with this new set of accessions.

4.2 Functional characterization of *FT* haplotypes with natural variation in the photoperiod control regions of *FT* in the rosette leaves

4.2.1 Characterization in F1 hybrids

ASE of *FT* revealed a group of candidate *FT* alleles expressed at a relatively similar level compared to the *Ler-FT* allele in F1 hybrids exposed to LD photoperiod and ambient temperature conditions. *FT* haplotypes selected for the TEPs surrounding *Block C* and located downstream of the 3'UTR of *FT* belong to this group (**Figure 13**). Therefore, the TEs seem to not affect the induction of *FT* expression in response to inductive conditions. We could speculate that the TEPs could have a paramutation effect on the *FT-Ler* allele. In

that particular context, the ASE assay in F1 hybrids could be inappropriate. However, paramutation interactions remain unclear in *Arabidopsis* (Xue et al. 2012). We could also consider the hypothesis that the TEs carry CREs increasing the induction of *FT* expression in inductive conditions. It would be interesting to evaluate the ASE profile of *FT* with the closest *FT* haplotypes to the TEP *FT* natural variants.

The screen for ASE of *FT* also showed the *FT-Est-1* mRNA to be slightly less abundant than the *Ler-FT* mRNA in F1 hybrids exposed to LD photoperiod and ambient temperature conditions (**Figure 13**). This result is unexpected since the *FT-Est-1* haplotype has been previously shown by QTL mapping with to delay the floral transition in an experimental population exposed to LDs in greenhouse conditions (Schwartz et al. 2009). This observation could be explained either by a floral transition highly sensitive to *FT* mRNA level or by the conditions used in our experiments. For example, diurnal temperature changes have been reported to alter *FT* expression profile (Kinmonth-Schultz et al. 2016). In that case, despite no significant differences in ASE profile in the timeframe used in my assays, I would not be able to evaluate the effect of the *FT* haplotypes accurately.

Taking the ASE profile of the *FT-Est-1* allele as a reference, a group of *FT* haplotypes turned to be weaker than the *FT-Ler* allele in response to inductive conditions, including LD photoperiod and ambient temperature. This group encompasses *FT* haplotypes selected for the presence of SNP in putative and functional TFBSs located at *Blocks A* and *C* (*REalpha*, *Shadow3A*, and *E-box*) (**Figure 13**).

ASE of *FT* expression in the rosette leaves of F1 hybrids revealed a group of *FT* alleles containing a SNP in the functional *CCAAT*-box located at *Block C* relatively insensitive to inductive conditions, including LD photoperiod and ambient temperature. Indeed, the *FT-Ler* allele showed a quasi mono-allelic expression in F1 hybrids in both inductive conditions (**Figure 13**). Consistent with previous reports, this *FT* haplotype has been proposed to be the causal polymorphism in the flowering time QTL mapped to the *FT-Ull2-5* allele (Strange et al. 2011).

4.2.2 Complementation experiments

The complementation of the *ft-10* mutant phenotype with the different constructs revealed a high variation of flowering time in the transgenic lines generated (**Figure 15**). A qPCR-based approach identified differences in transgene copy numbers in a population of T3 transgenic lines despite a segregation following a ratio of 3:1 in the previous T2 generation (**Figure 16A**). This transgene copy number variation affects my complementation assays dramatically. For example, the ability of the *TFÄ08p::FTgDNA* and *TFÄ08(funct)p::FTgDNA* constructs to complement the *ft-10* mutant phenotype is difficult to assess in this context (**Figure 16**). A screen for the presence of the transgene as a single copy is essential to only estimate the effect of the position of the transgene and ultimately characterize the functionality of the natural variation.

However, after screening for the presence of the transgene as a single copy, the *FÄLp::FTgDNA* and *Ale-stenar-56-14p::FTgDNA* constructs seem to be impaired in their ability to complement the late flowering phenotype of the *ft-10* mutant compared to the *Col-0p::FTgDNA* construct (**Figure 16**). These results suggest that the genetic variation within the 5,7 kb *FT* promoter of the FÄL and Ale-stenar-56-14 accessions, reduce the effect of *FT* during photoperiodic flowering. It is, for the moment, difficult to evaluate the effect of the SNP in the functional *CCAAT*-box located at *Block C* by complementation experiments. Indeed, *ft-10* plants transformed with the *Ale-stenar-56-14p::FTgDNA* and the *Ale-stenar-56-14(funct)p::FTgDNA* constructs have a comparable total leaf number before flowering (**Figure 16**). However, a CRISPR/Cas9 mediated SNP in the functional *CCAAT*-box located at *Block C* in a Col-0 background strongly affects the expression of *FT* and the floral transition in LD inductive conditions (Tingting Yang, unpublished data).

Taken together, the data support the causality of the SNP in the functional *CCAAT*-box located at *Block C* for the relatively insensitive *FT* haplotypes observed in the accessions Ale-stenar-56-14, UII2-5 and TV-10. The CO/NF-Y complex involved in the photoperiodic induction of *FT* expression is more likely not to bind the natural variant of the functional *CCAAT*-box located at *Block C*. NF-Y transcription factors have been proposed in mammals to act as pioneer factors by facilitating a permissive chromatin conformation (Fleming et al. 2013; Oldfield et al. 2014). Therefore, further investigations are needed to

characterize the chromatin state of *Block C* in the *FT* haplotype carrying a natural *CCAAT*-box variant.

4.3 The photoperiod and thermosensory induction of *FT* expression share CREs

Previous reports indicate induction of *FT* expression in the rosette leaves at dusk under SDs at 27°C (Balasubramanian et al. 2006). This thermosensory induction of *FT* expression has been shown to require PIF4, PIF5 and CO proteins (Balasubramanian et al. 2006; Kumar and Wigge 2010; Fernández et al. 2016). Since the CO mediated induction of *FT* expression requires the photoperiod control regions, *Blocks A* and *C*, we expected the candidate *FT* haplotypes to react in a similar manner in SDs at 27°C compared to photoperiod inductive conditions in F1 hybrids. Consistent with this assumption, the ASE profile of *FT* in the rosette leaves of F1 hybrids exposed to SDs at 27°C was comparable to photoperiod inductive conditions. Furthermore, this trend was also true for the relatively photoperiod insensitive *FT* haplotypes containing a SNP in the functional *CCAAT*-box located at *Block C*, since the *FT-Ler* allele was expressed with a quasi mono-allelic expression in F1 hybrids (**Figure 13**). This result suggests that the functional *CCAAT*-box located at *Block C*, specific to the NF-Y/CO complex, would be required in this thermosensory process.

Taken together, the data indicate that the photoperiod and thermosensory induction of *FT* expression share CREs. It would be interesting to investigate further the CREs required for the induction of *FT* expression in response to ambient temperature. The functional *CCAAT*-box located at *Block C* seem to be a good candidate already.

4.4 *FT* haplotypes less sensitive to inductive conditions are affected in a tissue specific manner

The regulation of *FT* expression in the siliques has been shown to be independent of *Block C* and to influence the floral reversion (Chen et al. 2014; Liu et al. 2014b). Interestingly, all candidate *FT* allele tested in F1 hybrids show a biallelic *FT* expression in developing siliques, including the alleles that were insensitive to inductive cues in the rosette leaves (**Figure 14**). This observation suggests that the CREs required for the induction of *FT*

Discussion

expression in the siliques are not altered in the candidate *FT* haplotypes. This result brings to question the CREs involved for the induction of *FT* expression in the siliques. *Block B*, phylogenetically conserved and located within the minimal promoter for the expression of *FT* in the siliques, seems to be a good candidate (Adrian et al. 2010). Additionally, preliminary results suggested an induction of *FT* expression in the siliques at ZT16 in LDs (Fabian Bratzel, unpublished data). Since *Block C* is not required for the induction of *FT* expression in the siliques, it would be interesting to know whether CO remains the main transcription factor involved in this process (Liu et al. 2014b).

Taken together, the ASE of *FT* in F1 hybrids indicates no difference of *FT* mRNA transcript level in the silique related developmental processes orchestrated by *FT*. The data suggest that the *FT* haplotypes identified in this study, which are less sensitive to inductive conditions, including photoperiod and ambient temperature, would be specifically affected in the floral transition. This observation is also remarkably true for the *FT* haplotype containing a SNP in the *CCAAT*-box located at *Block C*.

4.5 Footprint of relaxed selection of *FT* in accessions with a slow vernalization behavior?

The combination of ASE of *FT* in the rosette leaves of F1 hybrids exposed to LD photoperiod or ambient temperature revealed the *FT* haplotype carrying a SNP in the functional *CCAAT*-box located at *Block C* as the most altered *FT* haplotype in my assay. While this *FT* haplotype is relatively insensitive to inductive conditions in the rosette leaves, the induction of *FT* in developing siliques seem not to be affected, indicating that this *FT* haplotype is functionally differentiated in the floral transition specifically. This *FT* haplotype is present in accessions coming from the South of Sweden. Among these accessions is Ull2-5, an accession previously reported to have an *FLC* haplotype leading to an unusually slow vernalization behavior (Strange et al. 2011; Li et al. 2014).

As suggested in (Strange et al. 2011), we could expect that this *FT* haplotype would delay the floral transition, and therefore extend the vegetative development leading to more robust plants with higher seed yield. Natural *CCAAT*-box variants would provide a fitness advantage and would, consequently, be under positive selection. However, overwintered

Arabidopsis accessions in the North of Sweden have been reported to bolt within two weeks upon snowmelt (Duncan et al. 2015). Additionally, field experiments demonstrated a full vernalization response after 12 weeks of growth preceding winter for most of the accessions tested, including the Ull2-5 accession (Duncan et al. 2015). Consistent with the data, most of the candidate accessions selected in my screen, flowered with the same number of leaves compared to *Col-FRI_{s2}* plants in LDs, after a previous exposure to cold temperature in SDs during 12 weeks. However, the total number of leaves until flowering of *ft* null allele mutants is surprisingly comparable to wild type plants (**Figure 17**). Therefore, the *FT* haplotypes containing a natural *CCAAT*-box variant would not dramatically affect the flowering performance at least in the Ull2-5 background. In this evolutionary scenario, natural populations with slow vernalization behavior could tolerate less responsive *FT* haplotypes without affecting the pleiotropic effect of *FT* including floral reversion, seed production and ultimately fitness. Therefore, the photoperiod control regions of *FT* involved in the rosette leaves would be exposed to relaxed purifying selection.

4.6 Florigen has little impact on the floral transition after extensive periods of cold temperature

As the *FT* protein has been shown to be essential in the photoperiodic signal transmitted from the leaves to the SAM, the surprising low total number of leaves until flowering of *ft* null allele mutants brought into question the impact of the florigen on the floral transition after extensive periods of cold temperature. In a vernalization time course assay, the number of leaves in *ft-10* single mutants and *ft-10 tsf-1* double mutants compared to the wild-type plants indicated that florigen has less impact on the floral transition already after six weeks of cold temperature exposure. Consistent with the previously reported role of the florigen on the bolting time in ecologically realistic conditions (Brachi et al. 2013), the bolting time of *ft-10* single mutants and *ft-10 tsf-1* double mutants after a shift to LDs preceding prolonged exposure to cold temperature remains significantly different compared to wild-type plants. (**Figure 18**). Altogether, while *FT* influences total leaf number and bolting time in plants exposed to LD conditions, the effect of *FT* in plants previously exposed to long periods of cold temperature become specific to the bolting time. Therefore, the data suggests that another genetic pathway influences the total leaf number

Discussion

after extensive periods of cold temperature, thus bypassing the lack of florigen in term of flowering time.

Taken together, the data raises the question which genetic pathway is involved in response to this prolonged cold temperature exposure. Since the genetic background of the *ft-10* and *tsf-1* mutants contains a nonfunctional *FRI* allele, the floral induction in response to extensive periods of cold temperature is independent of *FRI*. *SOC1*, mainly expressed in the SAM, is likely to be involved in this process (Lee et al. 2000). However, the competence to flower is of course equally questionable. The age-related genetic systems, including the miR156/SPL and miR172/AP2-likes modules, contribute to the induction of the floral transition (Hyun et al. 2017). This finding could explain the rapid floral transition after snowmelt observed by colleagues (Duncan et al. 2015). However, another question remains. Why is the *FT* locus so conserved within the Brassicaceae family if its impact is strongly reduced after extended exposure to cold temperature? As mentioned before, the pleiotropic role of *FT* remains a possible explanation. However, the CREs involved in the induction of *FT* expression in the rosette leaves, including *Block C*, remain highly conserved.

4.7 The detection of functionally differentiated *FT* haplotypes is a complex scenario

The current methods to detect functionally differentiated haplotypes, including GWAS and QTL mappings, depend mainly on two components, the sample size corresponding to the allele frequency and the effect size of the phenotype (Korte and Farlow 2013; Weigel and Nordborg 2015). In the case of *FT*, the detection of functionally differentiated *FT* haplotypes is probably a complex scenario. Indeed, while exploring the whole spectrum of naturally occurring genetic variation at *FT* that could significantly alter its impact and potentially drive adaptation in *Arabidopsis* populations, I noticed that most of the *FT* haplotypes identified are rare in a global population of 1135 accessions. Therefore, the detection of functionally diverged *FT* haplotypes with the available tools would be limited by the sample size. Nevertheless, the relatively photoperiod insensitive *FT* haplotype carrying a SNP in the functional *CCAAT*-box located at *Block C* is the most functionally altered *FT* in my assay and represent the *FT* haplotype with the potentially largest effect.

However, at least in the case of the Ull2-5 accession background, the *FLC* allele, reported to have a distinctly slow vernalization behavior, would be epistatic to the *FT* allele prior cold exposure, masking the effect of the functionally distinguished *FT* haplotype. We can expect that due to the slow vernalization behavior provided by this particular *FLC* allele, the full vernalization would occur when *FT* has a weaker impact on the floral transition. Therefore, the effect size of this *FT* haplotype containing a SNP in the functional *CCAAT*-box located at *Block C* would be lower, thus limiting its detection in GWAS mapping. In the flowering time GWAS experiments done by the 1001 genomes consortium in a global population grown at 10°C and 16°C, we can notice that the peak of association at *FT* was only significant at 10°C (The 1001 Genomes Consortium 2016). We can speculate that at 16°C, the peak of association is low due to the epistatic interaction between *FLC* and *FT*. However, at 10°C the effect of *FLC* become weaker, revealing the effect of the functionally differentiated *FT* haplotypes. The best way to identify these *FT* haplotypes would be to use an experimental population to separate the epistatically acting *FLC* gene from *FT* in QTL mapping.

5. Material and methods

5.1 Screening for natural variation at *FT*

5.1.1 Single nucleotide polymorphisms

VCF files corresponding to the regions analyzed come from subset VCF, a tool provided by the 1001 genomes project gathering SNPs information of 1135 Arabidopsis accessions in a single file (<http://tools.1001genomes.org/vcfsubset/>). For SNPs located in coding sequences, snpEff VCF files provided information on the predicted effects of the corresponding SNP.

5.1.2 Structural variations

5.1.2.1 Mapping

Sequences from the 1001 genomes project are available online on NCBI in the SRA section (**Supplementary Table 5**). Paired-end short reads were mapped to the reference genome TAIR10 with BWA version 0.5.9-r16 (Li and Durbin 2009) and SAMtools version 0.1.19-44428cd (Li et al. 2009). The mapping process was executed with the Load Sharing Facility (LSF) scheduling platform on a GNU/Linux Debian 9 system.

```
% bwa index ref.fas
% fastq-dump --split-files $name.sra
% bwa aln -t 4 ref.fas $name_1.fastq > $name_1.sai
% bwa aln -t 4 ref.fas $name_2.fastq > $name_2.sai
% bwa sampe ref.fas $name_1.sai $name_2.sai $name_1.fastq$name_2.fastq >
$name.sam
% samtools view -bS $name.sam > $name.bam
% samtools sort $name.sorted.bam
% samtools index $name.sorted.bam
```

5.1.2.2 Structural variations detection

General

Pindel, version 0.2.5a8 (Ye et al. 2009) was used to detect structural variation from the bam files.

```
% pindel -f ref.fas -p $name.bam -c chrX:start-end -o output
```

The short insertion and deletion Pindel outputs (*_SI and *_D respectively) were converted to VCF files with pindel2vcf for downstream analysis.

```
% pindel2vcf -p output -r ref.fas -R TAIR10 -d date
```

Structural variations in coding sequences were further processed with SnpEff version 4.1i (build 2015-08-14) to predict the effect of the variant on the gene (Cingolani et al. 2012). The precompiled athalianaTair10 database was used for the analysis.

```
% java -Xm4g -jar snpEff.jar athalianaTair10 pindel2vcf_output.vcf > output
```

The discordant read-pair Pindel outputs (*_RP) were analyzed with a custom Python script to detect the TE insertions. Discordant read-pairs mapped to TEs annotated in the reference accessions Col-0 were used for the TE family prediction. Predicted TE insertions were confirmed by PCR on genomic DNA using oligonucleotide primers included in the Supplementary Table 4.

Functional characterization

The presence of reads was confirmed using sequences from Pseudogenomes, a tool provided by the 1001 genomes project gathering genomic sequences from 1135 Arabidopsis accessions in a single file. Putative structural variations were checked with Pindel. Structural variation predictions at *FT* were compared with a set of Arabidopsis accessions genotyped by Dr Liangyu Liu.

5.1.3 Phylogenetic footprinting

The sequence used in this study for the *FT* locus from the reference accession Col-0 is available on NCBI (<http://www.ncbi.nlm.nih.gov>). *FT* sequences of *Laevenworthia alabamica*, *Thellungiella halophila*, *Capsella rubella*, *Brassica rapa*, *Brassica stricta*, *Brassica napus*, *Arabis alpina*, *Arabidopsis halleri* and *Arabidopsis lyrata* used in this study were kindly provided by Dr Franziska Turck.

(<https://phytozome.jgi.doe.gov/pz/portal.html>).

mVista Shuffle-LAGAN was used for the large scale multiple alignments of gDNA from the different Brassicaceae species used in this study (Brudno et al. 2003; Frazer et al. 2004).

The resulting conserved sequences were then further aligned using MUSCLE (Edgar 2004).

The shadings were done with BOXSHADE.

(http://www.ch.embnet.org/software/BOX_form.html).

5.1.4 Screening for candidate accessions

Highly conserved nucleotides located in the putative and functional TFBSs were checked for SNPs and structural variations with a custom Python script. Accessions with natural variation within the coding sequence of *FT* were extracted based on the snpEff predictions and filtered for “HIGH” or “MODERATE” impacts with a custom python script. Accessions with natural variation within *cis*-elements of interest were extracted with a custom python script.

5.2 Plant Growth

Stratification was done at 4°C in SDs (8 h light/16 h dark) in cold chamber conditions (Celltherm) for four to five days to synchronize germination.

For crossing purposes and flowering time experiments after cold treatments, plants were first grown on soil at 21°C in SDs (8 h light/16 h dark) in growth chamber conditions (Percival, CLF) for two weeks, then transferred to 4°C in SDs in cold chamber conditions (Celltherm) for six or twelve weeks, depending on the experiment, and finally shifted to 21°C in LDs (16 h light/8 h dark) in growth chamber conditions (Percival, CLF).

For complementation experiments, transgenic lines were grown on soil at 21°C in LDs (16 h light/8 h dark) in growth chamber (Percival, CLF) or greenhouse conditions.

For ASE experiments, F1 hybrids were first grown on soil at 21°C in SDs (8 h light/16 h dark) in growth chamber conditions (Percival, CLF) for two weeks, transferred to 4°C in SDs in cold chamber conditions (Celltherm) for twelve weeks, moved to 21°C in SDs (8 h light/16 h dark) in growth chamber conditions (Percival, CLF) for two weeks, and finally shifted to 21°C in LDs (16 h light/8 h dark) or 27°C in SDs (8 h light/16 h dark) in growth chamber (Percival, CLF). One, three and five days after the shift, rosette leaves were harvested at ZT16 and ZT8 for the LD photoperiod and ambient temperature induction, respectively.

5.3 Allele-specific expression analysis

Ler was crossed with candidate accessions carrying a SNP (G/C) in the exon 4 of *FT* according to the method described in (Weigel and Glazebrook 2006a). F1 hybrids were confirmed by pyrosequencing with the corresponding gDNA following the procedure outlined below. Total mRNA and gDNA of two rosette leaves of F1 hybrids were extracted with TRIzol® Reagent (Life technologies™) and BioSprint 96 DNA (Qiagen®) respectively, following the manufacturer's protocol. Total mRNA of 12 to 16 developing siliques of F1 hybrids was extracted using a TRIzol®-based method (Meng and Feldman 2010). Two rosette leaves for at least six individuals were used for total mRNA and gDNA extraction. 12 to 16 developing siliques were used for mRNA extraction. 5 µg of mRNA was treated with DNaseI using the DNA-free™ kit (Ambion®) according to the manufacturer's protocol. Following reverse transcription using the Superscript II reverse transcriptase (Invitrogen™), *cDNA*, as well as gDNA, were amplified by polymerase chain reaction (PCR) using primers flanking the SNP (G/C) in the exon 4 of *FT* (**Supplementary Table 4**). This SNP allows differentiating *FT* alleles in F1 hybrids. The reverse primer is biotinylated to purify PCR products using streptavidin beads (GE Healthcare). The conditions of the PCR were: 94 °C for 2 min; 48 cycles of 94 °C for 45 s; 58°C for 45 s; 72°C for 45 s; and finally 72°C for 10 min. Amplicons were transferred with the sequencing primer (**Supplementary Table 4**) and sequenced with the PyroMark®Q96 ID (Qiagen®) and following the manufacturer's protocol. The sequencing results were analyzed with the PSQ™96MA (Qiagen®) software. Calibration of the pyrosequencing was done with gDNA extracted from two rosette leaves of two different accessions (Col-0 and HR-5) with BioSprint® 96 DNA (Qiagen®), following the manufacturer's protocol. These two

accessions were selected for their divergence at the SNP (G/C) in the exon 4 of *FT*. A dilution series of the two samples was performed to estimate the ratio between the concentrations by quantitative real-time PCR as described previously. gDNA from the two parents was then mixed in different ratios and determined using pyrosequencing. gDNA was included in all subsequent pyrosequencing experiments for underestimation correction and crossing rate.

5.4 Generation of transgenic lines

5.4.1 PCR

PCRs were performed using the Phusion® High-Fidelity DNA polymerase (New England Biolabs®) according to the manufacturer's protocol. The standard PCR reactions were following these conditions: 98 °C for 5 min; 30 cycles of 98 °C for 1 min; 58°C for 30 s; 72°C for 3 min; and finally 72°C for 10min. The PCR fragments were purified from agarose gel using the Nucleospin® Gel and PCR Clean-Up kit (Macherey-Nagel) according to the manufacturer's protocol.

5.4.2 Transformation of *Escherichia coli*

Transformation of plasmid DNA into *Escherichia coli* (*E.coli*) was done using the heat shock method (Froger and Hall 2007). The DH5α™ (Invitrogen™) *E.coli* strain was used for most of the plasmids, except for plasmids carrying the *ccdB* gene, transformed with DB3.1™ *E.coli* strain (Invitrogen™). After propagation in *E.coli* liquid culture with the appropriate antibiotics, the plasmids were extracted using Nucleospin® Plasmid kit (Macherey-Nagel) according to the manufacturer's protocol.

5.4.3 Plasmid construction

5.4.3.1 Generation of Gateway® ENTRY vectors

FT promoter variants were PCR-amplified from gDNA of Arabidopsis accessions (Col-0, Ale-stenar-56-14, TFÄ08, FäL1) using two different sets of primers (T855+T856 and

T854+T857) (**Supplementary Table 4**). In order to incorporate the natural *FT* promoter variants in a Gateway® *pENTR* vector, a *pENTR201* based plasmid was PCR-amplified using the primers T858 and T859 (**Supplementary Table 4**). The PCR products were combined, following the Golden Gate Assembly (New England Biolabs®) procedure (Engler et al. 2008), resulting in the *Col-0-5,7kbFTp-pENTR201*, *Ale-Stenar-56-14-5,7kbFTp-pENTR201*, *TFÄ08-5,7kbFTp-pENTR201* and *Fäl1-5,7kbFTp-pENTR201* entry vectors. The *FT* promoter variants introduced in the *pENTR201* construct were checked by Sanger sequencing.

Point mutations were induced in the *TFÄ08-5,7kbFTp-pENTR201* and *Ale-Stenar-56-14-5,7kbFTp-pENTR201* constructs by using two different sets of overlapping primers (T083+T084 and T087+T088, respectively) (**Supplementary Table 4**). Prior transformation in *E.coli*, the resulting *TFÄ08(funct)-5,7kbFTp-pENTR201* and *Ale-Stenar-56-14-(funct)-5,7kbFTp-pENTR201* constructs were distinguished from the original plasmids by DpnI (New England Biolabs®) digestion following the manufacturer's protocol. The PCR-induced point mutations in the *TFÄ08(funct)-5,7kbFTp-pENTR201* and *Ale-Stenar-56-14-5,7kbFTp-pENTR201* constructs were confirmed by Sanger sequencing.

5.4.3.2 Generation of Gateway® destination vectors with a selectable marker

In order to use the Fluorescent-accumulating seed technology (FAST) marker to select transformed and homozygous seeds, we took advantage of the *pFAST-G01* construct carrying the *GREEN FLUORESCENT PROTEIN (GFP)* driven by the *OLEOSIN 1 (OLE1)* promoter (Shimada et al. 2010). To incorporate the *OLE1p-GFP* cassette into a *pGreen0229* based vector, the *pFAST-G01* and the *GW-MCS-NOS-pGREEN* vectors were PCR-amplified with two different sets of primers (T026+T027 and T024+T025, respectively) (**Supplementary Table 4**). The subsequent PCR products were combined following the Gibson Assembly® (New England Biolabs) procedure, resulting to the *GW-MCS-NOS-pGREEN-OLE1* construct (Gibson et al. 2009). To incorporate an *FTgDNA* fragment, starting from the start codon to 2.3 kb downstream of the stop codon, into the *GW-MCS-NOS-pGREEN-OLE1* construct, an *FTgDNA-2.3kb* cassette was excised from the *GW-FTgDNA-2.3kb-pGREEN* construct, kindly provided by Dr. Tingting Ning, using the *XmaI* and *EcoNI* restriction enzymes (New England Biolabs) and ligated to the *GW-*

Material and methods

MCS-NOS-pGREEN-OLE1 construct prepared with the same restriction enzymes. The ligation product corresponds to the *GW-FTgDNA-2.3kb-pGREEN-OLE1* destination vector.

5.4.3.3 Generation of binary vectors

pEntry clones carrying the different natural *FT* promoter variants were recombined to the *GW-FTgDNA-2.3kb-pGREEN-OLE1* destination vector using the LR reaction (Invitrogen™) following the manufacturer's protocol. The resulting binary vectors correspond to *Col-0-5,7kbFTp-FTgDNA-2.3kb-pGREEN-OLE1*, *Ale-Stenar-56-14-5,7kbFTp-FTgDNA-pGREEN-OLE1*, *TFÄ08-5,7kbFTp-FTgDNA-pGREEN-OLE1*, *Fäl1-5,7kbFTp-FTgDNA-pGREEN-OLE1*, *TFÄ08(funct)-5,7kbFTp-FTgDNA-pGREEN-OLE1*, *Ale-Stenar-56-14(funct)-5,7kbFTp-FTgDNA-pGREEN-OLE1*.

5.4.4 Plant transformation and transgenic lines

All constructs based on the binary vector *pGreen0229* were introduced by electroporation (Weigel and Glazebrook 2006b) into *Agrobacterium tumefaciens* (Agrobacterium) strain GV3101, carrying the *pSOUP* helper plasmid. The transformed Agrobacterium were selected with the appropriate antibiotics according to standard lab procedures. Stable transgenic Agrobacterium lines were grown until OD600 0,6 – 1,2. Growth media was then supplemented with a solution composed of Sucrose 20% and Silvet L-77 [500 µL/L] in a 1:1 ratio. Constructs were transformed into *ft-10* mutants according to the floral dip method (Clough and Bent 1998). The *ft-10* mutant is a T-DNA insertion line in the Col-0 accession background (Yoo et al. 2005). The T-DNA is inserted in the first intron of *FT*. T1 generation was selected after germination on soil by spraying BASTA® [250 mg/L] (Bayer) (glufosinate-ammonium) three times with two days interval. Individuals with insertions at a single genomic location were selected in the T2 generation according to a segregation ratio between 2:1 and 4:1. Homozygous lines were identified in the T3 generation based on the GFP fluorescence provided by the *OLE1* cassette present in the inserted transgene.

5.4.5 Screening for tandem repeat insertions

gDNA of two rosette leaves of *ft-10* seedlings carrying the different *FT* promoter variant constructs were extracted with BioSprint 96 DNA (Qiagen®), following the manufacturer's protocol. Two set of primers (KK162, KK163 and KK164, KK165), kindly provided by Kristin Krause, were used for the quantification of *PP2A* and *PPT*, respectively. qPCRs were performed on a LightCycler480 (Roche) apparatus with a 10 µL PCR mix containing 1 µL of 10X Eva Buffer (0.7 mM dNTPs, 0.14 M Tris-HCl pH 8.8, 0.6 M KCl, and 36 mM MgCl₂), 0,5 µL of 20X EvaGreen® dye (Biotin), 0,1 µL of Taq polymerase (Biobudget), 0,25 µL of forward and reverse primers (10mM), 5,9 µL of H₂O and 2 µL of extracted gDNA. The conditions of the PCR were: 95 °C for 5 min; 40 cycles of 95 °C for 10 s; 60°C for 15 s; 72°C for 20 s; followed by a melting curve analysis of 95 °C for 5 s and 65 °C for 1 min, increasing from 65 °C to 97 °C with 2,5 °C/s. The amount of *PP2A* and *PPT* were calculated using a standard curve performed on a plasmid, kindly provided by Dr Franziska Turck, containing one copy of both *PP2A* and *PPT*. The amount of *PPT* relative to the housekeeping gene *PP2A* were calculated for three technical replicates. Control transgenic lines with a single copy of the transgene were previously genotyped by Southern Blot analysis

5.5 Flowering time measurement

Flowering time was scored by the number of rosette and cauline leaves on the main shoot of at least eight individuals before flowering. For the vernalization time course experiments, the bolting time (BT) and the total leaf number (TLN) were added. BT corresponded to the onset of elongation of the first internodes and was scored as the number of days from the germination required to distinguish the inflorescence from the leaves at a size < 5mm. TLN corresponded to the number of rosette and cauline leaves until flowering. The *ft-CRISPR* mutant lines were kindly provided by Dr. Youbong Hyun.

5.6 Statistical analysis

Statistical analyses were performed with the open source programming language R (R Core Team 2013). For flowering time scoring, ASE profile, and tandem repeat insertion screen, one-way ANOVA followed by a Holm-Sidak test ($P < 0.05$) were achieved with the “`pairwise.t.test`” function provided by the “`stat`” R package.

```
% anova = pairwise.t.test(dataframe$variable,dataframe$genotype, p.adjust =  
“holm”)
```

For the associations between *FT* haplotypes and geographical coordinates, the GLM was performed with the “`glm`” function provided by the “`stat`” R package.

```
%glm(formula = Lat ~ Admixture + FTvariant, data = dataframe)  
%glm(formula = Long ~ Admixture + FTvariant, data = dataframe)
```

5.7 Geographical distribution

The different geographical distributions were done using toner-background maps and the R package “`ggmap`” (Khale. D and Wickham. H 2013; R Core Team 2013), (<http://maps.stamen.com/#>), (<https://www.openstreetmap.org>).

References

- Abe M, Kobayashi Y, Yamamoto S, Daimon Y, Yamaguchi A, Ikeda Y, Ichinoki H, Notaguchi M, Goto K, Araki T. 2005. FD, a bZIP protein mediating signals from the floral pathway integrator FT at the shoot apex. *Science* **309**: 1052-1056.
- Adrian J, Farrona S, Reimer JJ, Albani MC, Coupland G, Turck F. 2010. cis-Regulatory elements and chromatin state coordinately control temporal and spatial expression of FLOWERING LOCUS T in Arabidopsis. *The Plant cell* **22**: 1425-1440.
- Agren J, Oakley CG, Lundemo S, Schemske DW. 2017. Adaptive divergence in flowering time among natural populations of Arabidopsis thaliana: Estimates of selection and QTL mapping. *Evolution; international journal of organic evolution* **71**: 550-564.
- Ahn JH, Miller D, Winter VJ, Banfield MJ, Lee JH, Yoo SY, Henz SR, Brady RL, Weigel D. 2006. A divergent external loop confers antagonistic activity on floral regulators FT and TFL1. *The EMBO journal* **25**: 605-614.
- Angel A, Song J, Dean C, Howard M. 2011. A Polycomb-based switch underlying quantitative epigenetic memory. *Nature* **476**: 105-108.
- Aukerman MJ, Sakai H. 2003. Regulation of flowering time and floral organ identity by a MicroRNA and its APETALA2-like target genes. *The Plant cell* **15**: 2730-2741.
- Ausin I, Alonso-Blanco C, Jarillo JA, Ruiz-Garcia L, Martinez-Zapater JM. 2004. Regulation of flowering time by FVE, a retinoblastoma-associated protein. *Nature genetics* **36**: 162-166.
- Balasubramanian S, Sureshkumar S, Lempe J, Weigel D. 2006. Potent induction of Arabidopsis thaliana flowering by elevated growth temperature. *PLoS genetics* **2**: e106.
- Bentsink L, Jowett J, Hanhart CJ, Koornneef M. 2006. Cloning of DOG1, a quantitative trait locus controlling seed dormancy in Arabidopsis. *Proceedings of the National Academy of Sciences of the United States of America* **103**: 17042-17047.
- Bergelson J, Roux F. 2010. Towards identifying genes underlying ecologically relevant traits in Arabidopsis thaliana. *Nature reviews Genetics* **11**: 867-879.
- Bergonzi S, Albani MC. 2011. Reproductive competence from an annual and a perennial perspective. *Journal of experimental botany* **62**: 4415-4422.
- Brachi B, Faure N, Bergelson J, Cuguen J, Roux F. 2013. Genome-wide association mapping of flowering time in nature: genetics for underlying components and reaction norms across two successive years. *Acta Bot Gallica* **160**: 205-219.
- Bratzel F, Turck F. 2015. Molecular memories in the regulation of seasonal flowering: from competence to cessation. *Genome biology* **16**: 192.
- Britten RJ, Kohne DE. 1968. Repeated sequences in DNA. Hundreds of thousands of copies of DNA sequences have been incorporated into the genomes of higher organisms. *Science* **161**: 529-540.
- Brudno M, Malde S, Poliakov A, Do CB, Couronne O, Dubchak I, Batzoglou S. 2003. Glocal alignment: finding rearrangements during alignment. *Bioinformatics* **19 Suppl 1**: i54-62.
- Cao J, Schneeberger K, Ossowski S, Gunther T, Bender S, Fitz J, Koenig D, Lanz C, Stegle O, Lippert C et al. 2011. Whole-genome sequencing of multiple Arabidopsis thaliana populations. *Nature genetics* **43**: 956-963.
- Cao S, Kumimoto RW, Gnesutta N, Calogero AM, Mantovani R, Holt BF, 3rd. 2014. A Distal CCAAT/NUCLEAR FACTOR Y Complex Promotes Chromatin Looping at the FLOWERING LOCUS T Promoter and Regulates the Timing of Flowering in Arabidopsis. *The Plant cell*.

References

- Chen M, MacGregor DR, Dave A, Florance H, Moore K, Paszkiewicz K, Smirnoff N, Graham IA, Penfield S. 2014. Maternal temperature history activates Flowering Locus T in fruits to control progeny dormancy according to time of year. *Proceedings of the National Academy of Sciences of the United States of America* **111**: 18787-18792.
- Chiang GC, Barua D, Kramer EM, Amasino RM, Donohue K. 2009. Major flowering time gene, flowering locus C, regulates seed germination in *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences of the United States of America* **106**: 11661-11666.
- Chien JC, Sussex IM. 1996. Differential regulation of trichome formation on the adaxial and abaxial leaf surfaces by gibberellins and photoperiod in *Arabidopsis thaliana* (L.) Heynh. *Plant physiology* **111**: 1321-1328.
- Cingolani P, Platts A, Wang le L, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM. 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly (Austin)* **6**: 80-92.
- Clough SJ, Bent AF. 1998. Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *The Plant journal : for cell and molecular biology* **16**: 735-743.
- Consortium EP. 2012. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**: 57-74.
- Copenhaver GP, Nickel K, Kuromori T, Benito MI, Kaul S, Lin X, Bevan M, Murphy G, Harris B, Parnell LD et al. 1999. Genetic definition and sequence analysis of *Arabidopsis* centromeres. *Science* **286**: 2468-2474.
- Corbesier L, Vincent C, Jang S, Fornara F, Fan Q, Searle I, Giakountis A, Farrona S, Gissot L, Turnbull C et al. 2007. FT protein movement contributes to long-distance signaling in floral induction of *Arabidopsis*. *Science* **316**: 1030-1033.
- Crevillen P, Yang H, Cui X, Greeff C, Trick M, Qiu Q, Cao X, Dean C. 2014. Epigenetic reprogramming that prevents transgenerational inheritance of the vernalized state. *Nature*.
- de la Chaux N, Tsuchimatsu T, Shimizu KK, Wagner A. 2012. The predominantly selfing plant *Arabidopsis thaliana* experienced a recent reduction in transposable element abundance compared to its outcrossing relative *Arabidopsis lyrata*. *Mob DNA* **3**: 2.
- De Lucia F, Crevillen P, Jones AM, Greb T, Dean C. 2008. A PHD-polycomb repressive complex 2 triggers the epigenetic silencing of FLC during vernalization. *Proceedings of the National Academy of Sciences of the United States of America* **105**: 16831-16836.
- DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, del Angel G, Rivas MA, Hanna M et al. 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature genetics* **43**: 491-498.
- Dittmar EL, Oakley CG, Agren J, Schemske DW. 2014. Flowering time QTL in natural populations of *Arabidopsis thaliana* and implications for their adaptive value. *Molecular ecology* **23**: 4291-4303.
- Duncan S, Holm S, Questa J, Irwin J, Grant A, Dean C. 2015. Seasonal shift in timing of vernalization as an adaptation to extreme winter. *Elife* **4**.
- Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic acids research* **32**: 1792-1797.

- El-Din El-Assal S, Alonso-Blanco C, Peeters AJ, Raz V, Koornneef M. 2001. A QTL for flowering time in Arabidopsis reveals a novel allele of CRY2. *Nature genetics* **29**: 435-440.
- El-Din El-Assal S, Alonso-Blanco C, Peeters AJ, Wagemaker C, Weller JL, Koornneef M. 2003. The role of cryptochrome 2 in flowering in Arabidopsis. *Plant physiology* **133**: 1504-1516.
- Engler C, Kandzia R, Marillonnet S. 2008. A one pot, one step, precision cloning method with high throughput capability. *PloS one* **3**: e3647.
- Fernández V, Germany MPIfPBRC, Takahashi Y, Germany MPIfPBRC, Le Gourrierc J, Germany MPIfPBRC, Coupland G, Germany MPIfPBRC. 2016. Photoperiodic and thermosensory pathways interact through CONSTANS to promote flowering at high temperature under short days. *The Plant Journal* **86**: 426-440.
- Fleming JD, Pavesi G, Benatti P, Imbriano C, Mantovani R, Struhl K. 2013. NF-Y coassociates with FOS at promoters, enhancers, repetitive elements, and inactive chromatin regions, and is stereo-positioned with growth-controlling transcription factors. *Genome research* **23**: 1195-1209.
- Frazer KA, Pachter L, Poliakov A, Rubin EM, Dubchak I. 2004. VISTA: computational tools for comparative genomics. *Nucleic acids research* **32**: W273-279.
- Froger A, Hall JE. 2007. Transformation of plasmid DNA into E. coli using the heat shock method. *J Vis Exp*: 253.
- Fujita T, Fujii H. 2013. Efficient isolation of specific genomic regions and identification of associated proteins by engineered DNA-binding molecule-mediated chromatin immunoprecipitation (enChIP) using CRISPR. *Biochemical and biophysical research communications* **439**: 132-136.
- Garner WW, Allard HA. 1920. EFFECT OF THE RELATIVE LENGTH OF DAY AND NIGHT AND OTHER FACTORS OF THE ENVIRONMENT ON GROWTH AND REPRODUCTION IN PLANTS1. [http://dxdoiorg/101175/1520-0493\(1920\)4820CO;2](http://dxdoiorg/101175/1520-0493(1920)4820CO;2).
- Gazzani S, Gendall AR, Lister C, Dean C. 2003. Analysis of the Molecular Basis of Flowering Time Variation in Arabidopsis Accessions1[w]. in *Plant physiology*, pp. 1107-1114.
- Gelvin SB. 2003. Agrobacterium-mediated plant transformation: the biology behind the "gene-jockeying" tool. *Microbiol Mol Biol Rev* **67**: 16-37, table of contents.
- Gibson DG, Young L, Chuang RY, Venter JC, Hutchison CA, 3rd, Smith HO. 2009. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature methods* **6**: 343-345.
- Grillo MA, Li C, Hammond M, Wang L, Schemske DW. 2013. Genetic architecture of flowering time differentiation between locally adapted populations of Arabidopsis thaliana. *The New phytologist* **197**: 1321-1331.
- Heo JB, Sung S. 2011. Vernalization-mediated epigenetic silencing by a long intronic noncoding RNA. *Science* **331**: 76-79.
- Ho WW, Weigel D. 2014. Structural features determining flower-promoting activity of Arabidopsis FLOWERING LOCUS T. *The Plant cell* **26**: 552-564.
- Huang X, Ding J, Effgen S, Turck F, Koornneef M. 2013. Multiple loci and genetic interactions involving flowering time genes regulate stem branching among natural variants of Arabidopsis. *The New phytologist* **199**: 843-857.
- Huijser P, Schmid M. 2011. The control of developmental phase transitions in plants. *Development* **138**: 4117-4129.

References

- Hyun Y, Kim J, Cho SW, Choi Y, Kim JS, Coupland G. 2015. Site-directed mutagenesis in *Arabidopsis thaliana* using dividing tissue-targeted RGEN of the CRISPR/Cas system to generate heritable null alleles. *Planta* **241**: 271-284.
- Hyun Y, Richter R, Coupland G. 2017. Competence to Flower: Age-Controlled Sensitivity to Environmental Cues. *Plant physiology* **173**: 36-46.
- Hyun Y, Richter R, Vincent C, Martinez-Gallegos R, Porri A, Coupland G. 2016. Multi-layered Regulation of SPL15 and Cooperation with SOC1 Integrate Endogenous Flowering Pathways at the *Arabidopsis* Shoot Meristem. *Developmental cell* **37**: 254-266.
- Iqbal Z, Caccamo M, Turner I, Flicek P, McVean G. 2012. De novo assembly and genotyping of variants using colored de Bruijn graphs. *Nature genetics* **44**: 226-232.
- Jang S, Marchal V, Panigrahi KC, Wenkel S, Soppe W, Deng XW, Valverde F, Coupland G. 2008. *Arabidopsis* COP1 shapes the temporal pattern of CO accumulation conferring a photoperiodic flowering response. *The EMBO journal* **27**: 1277-1288.
- Jiao WB, Schneeberger K. 2017. The impact of third generation genomic technologies on plant genome assembly. *Current opinion in plant biology* **36**: 64-70.
- Johanson U, West J, Lister C, Michaels S, Amasino R, Dean C. 2000. Molecular analysis of FRIGIDA, a major determinant of natural variation in *Arabidopsis* flowering time. *Science* **290**: 344-347.
- Kardailsky I, Shukla VK, Ahn JH, Dagenais N, Christensen SK, Nguyen JT, Chory J, Harrison MJ, Weigel D. 1999. Activation tagging of the floral inducer FT. *Science* **286**: 1962-1965.
- Kawakatsu T, Huang SS, Jupe F, Sasaki E, Schmitz RJ, Urich MA, Castanon R, Nery JR, Barragan C, He Y et al. 2016. Epigenomic Diversity in a Global Collection of *Arabidopsis thaliana* Accessions. *Cell* **166**: 492-505.
- Khale. D, Wickham. H. 2013. ggmap: Spatial Visualization with ggplot2. *The R Journal* **5**: 144-161.
- Kim DH, Sung S. 2017. Vernalization-Triggered Intragenic Chromatin Loop Formation by Long Noncoding RNAs. *Developmental cell* **40**: 302-312.e304.
- Kinmonth-Schultz HA, Tong X, Lee J, Song YH, Ito S, Kim SH, Imaizumi T. 2016. Cool night-time temperatures induce the expression of CONSTANS and FLOWERING LOCUS T to regulate flowering in *Arabidopsis*. *The New phytologist*.
- Kinoshita T, Ono N, Hayashi Y, Morimoto S, Nakamura S, Soda M, Kato Y, Ohnishi M, Nakano T, Inoue S et al. 2011. FLOWERING LOCUS T regulates stomatal opening. *Current biology : CB* **21**: 1232-1238.
- Klasen JR, Barbez E, Meier L, Meinshausen N, Bühlmann P, Koornneef M, Busch W, Schneeberger K. 2016. A multi-marker association method for genome-wide association studies without the need for population structure correction. *Nature Communications, Published online: 10 November 2016; | doi:101038/ncomms13299*.
- Kobayashi Y, Kaya H, Goto K, Iwabuchi M, Araki T. 1999. A pair of related genes with antagonistic roles in mediating flowering signals. *Science* **286**: 1960-1962.
- Koini MA, Alvey L, Allen T, Tilley CA, Harberd NP, Whitelam GC, Franklin KA. 2009. High temperature-mediated adaptations in plant architecture require the bHLH transcription factor PIF4. *Current biology : CB* **19**: 408-413.
- Koornneef M, Alonso-Blanco C, Vreugdenhil D. 2004. Naturally occurring genetic variation in *Arabidopsis thaliana*. *Annual review of plant biology* **55**: 141-172.

- Koornneef M, Hanhart CJ, van der Veen JH. 1991. A genetic and physiological analysis of late flowering mutants in *Arabidopsis thaliana*. *Mol Gen Genet* **229**: 57-66.
- Koornneef M, Meinke D. 2010. The development of *Arabidopsis* as a model plant. *The Plant journal : for cell and molecular biology* **61**: 909-921.
- Korte A, Farlow A. 2013. The advantages and limitations of trait analysis with GWAS: a review. in *Plant methods*, p. 29.
- Korves TM, Schmid KJ, Caicedo AL, Mays C, Stinchcombe JR, Purugganan MD, Schmitt J. 2007. Fitness effects associated with the major flowering time gene *FRIGIDA* in *Arabidopsis thaliana* in the field. *Am Nat* **169**: E141-157.
- Kotake T, Takada S, Nakahigashi K, Ohto M, Goto K. 2003. *Arabidopsis* *TERMINAL FLOWER 2* gene encodes a heterochromatin protein 1 homolog and represses both *FLOWERING LOCUS T* to regulate flowering time and several floral homeotic genes. *Plant & cell physiology* **44**: 555-564.
- Kumar SV, Lucyshyn D, Jaeger KE, Alós E, Alvey E, Harberd NP, Wigge PA. 2012. Transcription factor PIF4 controls the thermosensory activation of flowering. *Nature* **484**: 242-245.
- Kumar SV, Wigge PA. 2010. H2A.Z-containing nucleosomes mediate the thermosensory response in *Arabidopsis*. *Cell* **140**: 136-147.
- Kumimoto RW, Adam L, Hymus GJ, Repetti PP, Reuber TL, Marion CM, Hempel FD, Ratcliffe OJ. 2008. The Nuclear Factor Y subunits NF-YB2 and NF-YB3 play additive roles in the promotion of flowering by inductive long-day photoperiods in *Arabidopsis*. *Planta* **228**: 709-723.
- Lee H, Suh SS, Park E, Cho E, Ahn JH, Kim SG, Lee JS, Kwon YM, Lee I. 2000. The AGAMOUS-LIKE 20 MADS domain protein integrates floral inductive pathways in *Arabidopsis*. *Genes & development* **14**: 2366-2376.
- Lee JH, Ryu HS, Chung KS, Pose D, Kim S, Schmid M, Ahn JH. 2013. Regulation of temperature-responsive flowering by MADS-box transcription factor repressors. *Science* **342**: 628-632.
- Lee JH, Yoo SJ, Park SH, Hwang I, Lee JS, Ahn JH. 2007. Role of SVP in the control of flowering time by ambient temperature in *Arabidopsis*. *Genes & development* **21**: 397-402.
- Li D, Liu C, Shen L, Wu Y, Chen H, Robertson M, Helliwell CA, Ito T, Meyerowitz E, Yu H. 2008. A repressor complex governs the integration of flowering signals in *Arabidopsis*. *Developmental cell* **15**: 110-120.
- Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**: 1754-1760.
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**: 2078-2079.
- Li P, Filiault D, Box MS, Kerdaffrec E, van Oosterhout C, Wilczek AM, Schmitt J, McMullan M, Bergelson J, Nordborg M et al. 2014. Multiple FLC haplotypes defined by independent cis-regulatory variation underpin life history diversity in *Arabidopsis thaliana*. *Genes & development* **28**: 1635-1640.
- Li Y, Huang Y, Bergelson J, Nordborg M, Borevitz JO. 2010. Association mapping of local climate-sensitive quantitative trait loci in *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences of the United States of America* **107**: 21199-21204.
- Liu L, Adrian J, Pankin A, Hu J, Dong X, von Korff M, Turck F. 2014a. Induced and natural variation of promoter length modulates the photoperiodic response of *FLOWERING LOCUS T*. *Nature communications* **5**: 4558.

References

- Liu L, Farrona S, Klemme S, Turck FK. 2014b. Post-fertilization expression of FLOWERING LOCUS T suppresses reproductive reversion. *Frontiers in plant science* **5**: 164.
- Liu L, Liu C, Hou X, Xi W, Shen L, Tao Z, Wang Y, Yu H. 2012. FTIP1 is an essential regulator required for florigen transport. *PLoS biology* **10**: e1001313.
- Lutz U, Pose D, Pfeifer M, Gundlach H, Hagmann J, Wang C, Weigel D, Mayer KF, Schmid M, Schwechheimer C. 2015. Modulation of Ambient Temperature-Dependent Flowering in *Arabidopsis thaliana* by Natural Variation of FLOWERING LOCUS M. *PLoS genetics* **11**: e1005588.
- Marquardt S, Boss PK, Hadfield J, Dean C. 2006. Additional targets of the *Arabidopsis* autonomous pathway members, FCA and FY. *Journal of experimental botany* **57**: 3379-3386.
- Martinez-Zapater JM, Somerville CR. 1990. Effect of Light Quality and Vernalization on Late-Flowering Mutants of *Arabidopsis thaliana*. *Plant physiology* **92**: 770-776.
- Mathieu J, Yant LJ, Murdter F, Kuttner F, Schmid M. 2009. Repression of flowering by the miR172 target SMZ. *PLoS biology* **7**: e1000148.
- Meng L, Feldman L. 2010. A rapid TRIzol-based two-step method for DNA-free RNA extraction from *Arabidopsis* siliques and dry seeds. *Biotechnol J* **5**: 183-186.
- Michaels SD, Amasino RM. 1999. FLOWERING LOCUS C encodes a novel MADS domain protein that acts as a repressor of flowering. *The Plant cell* **11**: 949-956.
- Michaels SD, He Y, Scortecci KC, Amasino RM. 2003. Attenuation of FLOWERING LOCUS C activity as a mechanism for the evolution of summer-annual flowering behavior in *Arabidopsis*. *Proceedings of the National Academy of Sciences of the United States of America* **100**: 10102-10107.
- Naumova N, Smith EM, Zhan Y, Dekker J. 2012. Analysis of long-range chromatin interactions using Chromosome Conformation Capture. *Methods* **58**: 192-203.
- Oldfield AJ, Yang P, Conway AE, Cinghu S, Freudenberg JM, Yellaboina S, Jothi R. 2014. Histone-fold domain protein NF-Y promotes chromatin accessibility for cell type-specific master transcription factors. *Molecular cell* **55**: 708-722.
- Pastinen T. 2010. Genome-wide allele-specific analysis: insights into regulatory variation. *Nature reviews Genetics* **11**: 533-538.
- Pevzner PA, Tang H, Waterman MS. 2001. An Eulerian path approach to DNA fragment assembly. *Proceedings of the National Academy of Sciences of the United States of America* **98**: 9748-9753.
- Pin PA, Nilsson O. 2012. The multifaceted roles of FLOWERING LOCUS T in plant development. *Plant, cell & environment* **35**: 1742-1755.
- Porri A, Torti S, Romera-Branchat M, Coupland G. 2012. Spatially distinct regulatory roles for gibberellins in the promotion of flowering of *Arabidopsis* under long photoperiods. *Development* **139**: 2198-2209.
- Pose D, Verhage L, Ott F, Yant L, Mathieu J, Angenent GC, Immink RG, Schmid M. 2013. Temperature-dependent regulation of flowering by antagonistic FLM variants. *Nature*.
- Putterill J, Robson F, Lee K, Simon R, Coupland G. 1995. The CONSTANS gene of *Arabidopsis* promotes flowering and encodes a protein showing similarities to zinc finger transcription factors. *Cell* **80**: 847-857.
- Qüesta JI, Song J, Geraldo N, An H, Dean C. 2016. *Arabidopsis* transcriptional repressor VAL1 triggers Polycomb silencing at FLC during vernalization.
- R Core Team. 2013. R: A Language and Environment for Statistical Computing. in R Foundation for Statistical Computing, Vienna, Austria.

- Rhoads A, Au KF. 2015. PacBio Sequencing and Its Applications. *Genomics Proteomics Bioinformatics* **13**: 278-289.
- Robinson JT, Thorvaldsdottir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP. 2011. Integrative genomics viewer. *Nature biotechnology* **29**: 24-26.
- Romier C, Cocchiarella F, Mantovani R, Moras D. 2003. The NF-YB/NF-YC structure gives insight into DNA binding and transcription regulation by CCAAT factor NF-Y. *The Journal of biological chemistry* **278**: 1336-1345.
- Roux F, Laboratoire de Génétique et Evolution des Populations Végétales U-C, FR CNRS 1818, Université de Lille I, F-59655 Villeneuve d'Ascq Cedex, France, Present address: Department of Ecology and Evolution UoC, 1101 East 57th Street, Chicago, IL 60637, USA, Touzet P, Laboratoire de Génétique et Evolution des Populations Végétales U-C, FR CNRS 1818, Université de Lille I, F-59655 Villeneuve d'Ascq Cedex, France, Cuguen J, Laboratoire de Génétique et Evolution des Populations Végétales U-C, FR CNRS 1818, Université de Lille I, F-59655 Villeneuve d'Ascq Cedex, France, Corre VL, lecorre@dijon.inra.fr, UMR 1210 Biologie et Gestion des Adventices INdIRA, 17 rue Sully, F-21065 Dijon Cedex, France. 2006. How to be early flowering: an evolutionary perspective. *Trends in plant science* **11**: 375-381.
- Salomé PA, Bomblies K, Laitinen RAE, Yant L, Mott R, Weigel D. 2011. Genetic Architecture of Flowering-Time Variation in *Arabidopsis thaliana*. in *Genetics*, pp. 421-433.
- Samach A, Onouchi H, Gold SE, Ditta GS, Schwarz-Sommer Z, Yanofsky MF, Coupland G. 2000. Distinct roles of CONSTANS target genes in reproductive development of *Arabidopsis*. *Science* **288**: 1613-1616.
- Sanchez-Bermejo E, Balasubramanian S. 2015. Natural variation involving deletion alleles of FRIGIDA modulate temperature-sensitive flowering responses in *Arabidopsis thaliana*. *Plant, cell & environment*.
- Sati S, Cavalli G. 2017. Chromosome conformation capture technologies and their impact in understanding genome function. *Chromosoma* **126**: 33-44.
- Sawa M, Kay SA. 2011. GIGANTEA directly activates Flowering Locus T in *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences of the United States of America* **108**: 11698-11703.
- Sawa M, Nusinow DA, Kay SA, Imaizumi T. 2007. FKF1 and GIGANTEA complex formation is required for day-length measurement in *Arabidopsis*. *Science* **318**: 261-265.
- Schmalenbach I, Zhang L, Ryngajllo M, Jimenez-Gomez JM. 2014. Functional analysis of the Landsberg erecta allele of FRIGIDA. *BMC plant biology* **14**: 218.
- Schmid M, Uhlenhaut NH, Godard F, Demar M, Bressan R, Weigel D, Lohmann JU. 2003. Dissection of floral induction pathways using global expression analysis. *Development* **130**: 6001-6012.
- Schwartz C, Balasubramanian S, Warthmann N, Michael TP, Lempe J, Sureshkumar S, Kobayashi Y, Maloof JN, Borevitz JO, Chory J et al. 2009. Cis-regulatory changes at FLOWERING LOCUS T mediate natural variation in flowering responses of *Arabidopsis thaliana*. *Genetics* **183**: 723-732, 721SI-727SI.
- Scortecci KC, Michaels SD, Amasino RM. 2001. Identification of a MADS-box gene, FLOWERING LOCUS M, that represses flowering. *The Plant journal : for cell and molecular biology* **26**: 229-236.
- Searle I, He Y, Turck F, Vincent C, Fornara F, Krober S, Amasino RA, Coupland G. 2006. The transcription factor FLC confers a flowering response to vernalization

References

- by repressing meristem competence and systemic signaling in Arabidopsis. *Genes & development* **20**: 898-912.
- Seren U, Grimm D, Fitz J, Weigel D, Nordborg M, Borgwardt K, Korte A. 2017. AraPheno: a public database for Arabidopsis thaliana phenotypes. *Nucleic acids research* **45**: D1054-d1059.
- Sheldon CC, Hills MJ, Lister C, Dean C, Dennis ES, Peacock WJ. 2008. Resetting of FLOWERING LOCUS C expression after epigenetic repression by vernalization. *Proceedings of the National Academy of Sciences of the United States of America* **105**: 2214-2219.
- Shi Z, Maximova SN, Liu Y, Verica J, Guiltinan MJ. 2010. Functional analysis of the Theobroma cacao NPR1 gene in Arabidopsis. *BMC plant biology* **10**: 248.
- Shimada TL, Shimada T, Hara-Nishimura I. 2010. A rapid and non-destructive screenable marker, FAST, for identifying transformed seeds of Arabidopsis thaliana. *The Plant journal : for cell and molecular biology* **61**: 519-528.
- Shindo C, Aranzana MJ, Lister C, Baxter C, Nicholls C, Nordborg M, Dean C. 2005. Role of FRIGIDA and FLOWERING LOCUS C in determining variation in flowering time of Arabidopsis. *Plant physiology* **138**: 1163-1173.
- Siefers N, Dang KK, Kumimoto RW, Bynum WE, Tayrose G, Holt BF, 3rd. 2009. Tissue-specific expression patterns of Arabidopsis NF-Y transcription factors suggest potential for extensive combinatorial complexity. *Plant physiology* **149**: 625-641.
- Slotkin RK, Martienssen R. 2007. Transposable elements and the epigenetic regulation of the genome. *Nature Reviews Genetics* **8**: 272-285.
- Song YH, Smith RW, To BJ, Millar AJ, Imaizumi T. 2012a. FKF1 Conveys Timing Information for CONSTANS Stabilization in Photoperiodic Flowering.
- Song YH, Smith RW, To BJ, Millar AJ, Imaizumi T. 2012b. FKF1 conveys timing information for CONSTANS stabilization in photoperiodic flowering. *Science* **336**: 1045-1049.
- Strange A, Li P, Lister C, Anderson J, Warthmann N, Shindo C, Irwin J, Nordborg M, Dean C. 2011. Major-effect alleles at relatively few loci underlie distinct vernalization and flowering variation in Arabidopsis accessions. *PloS one* **6**: e19949.
- Swiezewski S, Liu F, Magusin A, Dean C. 2009. Cold-induced silencing by long antisense transcripts of an Arabidopsis Polycomb target. *Nature* **462**: 799-802.
- Takada S, Goto K. 2003. TERMINAL FLOWER2, an Arabidopsis Homolog of HETEROCHROMATIN PROTEIN1, Counteracts the Activation of FLOWERING LOCUS T by CONSTANS in the Vascular Tissues of Leaves to Regulate Flowering Time. in *The Plant cell*, pp. 2856-2865.
- Tattini L, D'Aurizio R, Magi A. 2015. Detection of Genomic Structural Variants from Next-Generation Sequencing Data. *Front Bioeng Biotechnol* **3**: 92.
- Telfer A, Bollman KM, Poethig RS. 1997. Phase change and the regulation of trichome distribution in Arabidopsis thaliana. *Development* **124**: 645-654.
- The 1001 Genomes Consortium. 2016. 1,135 Genomes Reveal the Global Pattern of Polymorphism in Arabidopsis thaliana. *Cell*.
- The Arabidopsis Genome Initiative. 2000. Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. *Nature* **408**: 796-815.
- Tiwari SB, Shen Y, Chang HC, Hou Y, Harris A, Ma SF, McPartland M, Hymus GJ, Adam L, Marion C et al. 2010. The flowering time regulator CONSTANS is recruited to the FLOWERING LOCUS T promoter via a unique cis-element. *The New phytologist* **187**: 57-66.

- Valverde F, Mouradov A, Soppe W, Ravenscroft D, Samach A, Coupland G. 2004. Photoreceptor regulation of CONSTANS protein in photoperiodic flowering. *Science* **303**: 1003-1006.
- Wang H, Ma LG, Li JM, Zhao HY, Deng XW. 2001. Direct interaction of Arabidopsis cryptochromes with COP1 in light control development. *Science* **294**: 154-158.
- Wang J-W, Department of Molecular Biology MPIfDB, D-72076 Tübingen, Germany, Czech B, Department of Molecular Biology MPIfDB, D-72076 Tübingen, Germany, Weigel D, weigel@weigelworld.org, Department of Molecular Biology MPIfDB, D-72076 Tübingen, Germany. 2009. miR156-Regulated SPL Transcription Factors Define an Endogenous Flowering Pathway in Arabidopsis thaliana. *Cell* **138**: 738-749.
- Weigel D. 2012. Natural variation in Arabidopsis: from molecular genetics to ecological genomics. *Plant physiology* **158**: 2-22.
- Weigel D, Glazebrook J. 2006a. Setting Up Arabidopsis Crosses.
- Weigel D, Glazebrook J. 2006b. Transformation of agrobacterium using electroporation. *CSH Protoc* **2006**.
- Weigel D, Nordborg M. 2015. Population Genomics for Understanding Adaptation in Wild Plant Species. *Annual review of genetics*.
- Wenkel S, Turck F, Singer K, Gissot L, Le Gourrierc J, Samach A, Coupland G. 2006. CONSTANS and the CCAAT box binding complex share a functionally important domain and interact to regulate flowering of Arabidopsis. *The Plant cell* **18**: 2971-2984.
- Werner JD, Borevitz JO, Uhlenhaut NH, Ecker JR, Chory J, Weigel D. 2005. FRIGIDA-independent variation in flowering time of natural Arabidopsis thaliana accessions. *Genetics* **170**: 1197-1207.
- Wilson RN, Heckman JW, Somerville CR. 1992. Gibberellin Is Required for Flowering in Arabidopsis thaliana under Short Days.
- Xue W, Ruprecht C, Street N, Hematy K, Chang C, Frommer WB, Persson S, Niittyla T. 2012. Paramutation-like interaction of T-DNA loci in Arabidopsis. *PloS one* **7**: e51651.
- Yamaguchi A, Kobayashi Y, Goto K, Abe M, Araki T. 2005. TWIN SISTER OF FT (TSF) acts as a floral pathway integrator redundantly with FT. *Plant & cell physiology* **46**: 1175-1189.
- Yamaguchi A, Wu MF, Yang L, Wu G, Poethig RS, Wagner D. 2009. The microRNA-regulated SBP-Box transcription factor SPL3 is a direct upstream activator of LEAFY, FRUITFULL, and APETALA1. *Developmental cell* **17**: 268-278.
- Yang HQ, Wu YJ, Tang RH, Liu D, Liu Y, Cashmore AR. 2000. The C termini of Arabidopsis cryptochromes mediate a constitutive light response. *Cell* **103**: 815-827.
- Ye K, Schulz MH, Long Q, Apweiler R, Ning Z. 2009. Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. *Bioinformatics* **25**: 2865-2871.
- Yoo SK, Chung KS, Kim J, Lee JH, Hong SM, Yoo SJ, Yoo SY, Lee JS, Ahn JH. 2005. CONSTANS activates SUPPRESSOR OF OVEREXPRESSION OF CONSTANS 1 through FLOWERING LOCUS T to promote flowering in Arabidopsis. *Plant physiology* **139**: 770-778.
- Yuan W, Luo X, Li Z, Yang W, Wang Y, Liu R, Du J, He Y. 2016. A cis cold memory element and a trans epigenome reader mediate Polycomb silencing of FLC by vernalization in Arabidopsis. *Nature genetics* **48**: 1527-1534.

References

- Zhang W, Zhang T, Wu Y, Jiang J. 2012. Genome-wide identification of regulatory DNA elements and protein-binding footprints using signatures of open chromatin in Arabidopsis. *The Plant cell* **24**: 2719-2731.
- Zuo Z, Liu H, Liu B, Liu X, Lin C. 2011. Blue Light-Dependent Interaction of CRY2 with SPA1 Regulates COP1 activity and Floral Initiation in Arabidopsis. *Current biology : CB* **21**: 841-847.

Appendix

A

```
Call:
glm(formula = Lat ~ Admixture + Coding, data = Df)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-116.5    -1.7     0.1     1.1   7366.5

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    151.3538    60.3106   2.510  0.01227 *
Admixturecentral_europe -103.4848    35.5237  -2.913  0.00367 **
Admixturegermany      -102.5473    40.0744  -2.559  0.01067 *
Admixtureitaly_balkan_caucasus -110.6546    39.8761  -2.775  0.00564 **
Admixturenorth_sweden   -88.8338    44.4008  -2.001  0.04574 *
Admixturerelict        -113.9329    58.8302  -1.937  0.05312 .
Admixturesouth_sweden   -95.4367    36.8052  -2.593  0.00968 **
Admixturespain         -110.5056    38.4643  -2.873  0.00417 **
Admixturewestern_europe -102.9261    39.1167  -2.631  0.00866 **
Coding+                0.1333     52.3961   0.003  0.99797
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for gaussian family taken to be 64224.61)

Null deviance: 55774461 on 866 degrees of freedom
Residual deviance: 55040492 on 857 degrees of freedom
AIC: 12070

Number of Fisher Scoring iterations: 2
```

B

```
Call:
glm(formula = Long ~ Admixture + Coding, data = Df)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-151.665   -1.963   -0.045    5.155   122.620

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    65.475338    4.771874   13.721 <2e-16 ***
Admixturecentral_europe -52.347724    2.810694  -18.624 <2e-16 ***
Admixturegermany      -87.711842    3.170751  -27.663 <2e-16 ***
Admixtureitaly_balkan_caucasus -35.517919    3.155064  -11.257 <2e-16 ***
Admixturenorth_sweden  -47.129192    3.513068  -13.415 <2e-16 ***
Admixturerelict       -71.281094    4.654746  -15.314 <2e-16 ***
Admixturesouth_sweden  -51.208091    2.912092  -17.585 <2e-16 ***
Admixturespain        -67.704082    3.043357  -22.247 <2e-16 ***
Admixturewestern_europe -71.314521    3.094976  -23.042 <2e-16 ***
Coding+           0.004244    4.145671   0.001  0.999
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for gaussian family taken to be 402.0617)

Null deviance: 751962 on 866 degrees of freedom
Residual deviance: 344567 on 857 degrees of freedom
AIC: 7671.4

Number of Fisher Scoring iterations: 2
```

Supplementary Figure 1: Geographical distribution of the candidate *FT* haplotypes selected for natural variation in the coding sequence of *FT*.

(A, B) Test of the latitudinal and longitudinal distributions of the candidate *FT* haplotypes selected for natural variation in the coding sequence of *FT* in (A) and (B), respectively, using a GLM model that includes an estimate of the population structure previously described (The 1001 Genomes Consortium 2016).

Appendix

A

```
Call:
glm(formula = Lat ~ Admixture + FTpromoter, data = Df)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-21.9361  -0.6944   0.0905   1.1438  14.0723

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    45.0977     0.6183  72.935 < 2e-16 ***
Admixturecentral_europe  2.8410     0.7120   3.990 7.60e-05 ***
Admixturegermany      3.9558     0.7652   5.169 3.40e-07 ***
Admixtureitaly_balkan_caucasus -3.8100     0.8331  -4.573 6.06e-06 ***
Admixturenorth_sweden  17.3691     0.8543  20.330 < 2e-16 ***
Admixturerelict       -8.0505     1.1126  -7.236 1.75e-12 ***
Admixturesouth_sweden  10.8718     0.7089  15.336 < 2e-16 ***
Admixturespain       -4.0617     0.7928  -5.123 4.30e-07 ***
Admixturewestern_europe  4.2036     0.7849   5.356 1.30e-07 ***
FTpromoterlong        0.7612     0.3843   1.981 0.0482 *
FTpromotershort      -0.6910     0.7765  -0.890 0.3739
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for gaussian family taken to be 12.50495)

Null deviance: 26340.3 on 510 degrees of freedom
Residual deviance: 6252.5 on 500 degrees of freedom
AIC: 2753.9

Number of Fisher Scoring iterations: 2
```

B

```
Call:
glm(formula = Lat ~ Admixture + FTpromoter, data = Df)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-21.9361  -0.6944   0.0905   1.1438  14.0723

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    45.8588     0.6966  65.831 < 2e-16 ***
Admixturecentral_europe  2.8410     0.7120   3.990 7.60e-05 ***
Admixturegermany      3.9558     0.7652   5.169 3.40e-07 ***
Admixtureitaly_balkan_caucasus -3.8100     0.8331  -4.573 6.06e-06 ***
Admixturenorth_sweden  17.3691     0.8543  20.330 < 2e-16 ***
Admixturerelict       -8.0505     1.1126  -7.236 1.75e-12 ***
Admixturesouth_sweden  10.8718     0.7089  15.336 < 2e-16 ***
Admixturespain       -4.0617     0.7928  -5.123 4.30e-07 ***
Admixturewestern_europe  4.2036     0.7849   5.356 1.30e-07 ***
FTpromotermedium     -0.7612     0.3843  -1.981 0.0482 *
FTpromotershort      -1.4522     0.8312  -1.747 0.0812 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for gaussian family taken to be 12.50495)

Null deviance: 26340.3 on 510 degrees of freedom
Residual deviance: 6252.5 on 500 degrees of freedom
AIC: 2753.9

Number of Fisher Scoring iterations: 2
```

Supplementary Figure 2: Latitudinal distribution of the three main *FT* promoter types.

(A, B) Test of the latitudinal distribution of the *FT* promoter types using a GLM model that includes an estimate of the population structure previously described (The 1001 Genomes Consortium 2016). The medium and long *FT* promoter types were used as a reference in (A) and (B), respectively.

A

```

Call:
glm(formula = Long ~ Admixture + FTpromoter, data = Df)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-150.606   -2.537    -0.163     7.237   124.050

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)      62.146      3.903   15.921 < 2e-16 ***
Admixturecentral_europe -50.444      4.495  -11.223 < 2e-16 ***
Admixturegermany     -79.147      4.831  -16.384 < 2e-16 ***
Admixtureitaly_balkan_caucasus -31.819      5.259   -6.051 2.83e-09 ***
Admixturenorth_sweden -43.578      5.393   -8.080 4.89e-15 ***
Admixturerelict      -69.634      7.023   -9.915 < 2e-16 ***
Admixturesouth_sweden -48.011      4.475  -10.729 < 2e-16 ***
Admixturespain       -63.829      5.005  -12.754 < 2e-16 ***
Admixturewestern_europe -70.021      4.955  -14.132 < 2e-16 ***
FTpromoterlong        2.274      2.426    0.937  0.349
FTpromotershort      -4.648      4.902   -0.948  0.343
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for gaussian family taken to be 498.3158)

Null deviance: 435215  on 510  degrees of freedom
Residual deviance: 249158  on 500  degrees of freedom
AIC: 4637

Number of Fisher Scoring iterations: 2

```

B

```

Call:
glm(formula = Long ~ Admixture + FTpromoter, data = Df)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-150.606   -2.537    -0.163     7.237   124.050

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)      64.420      4.397   14.649 < 2e-16 ***
Admixturecentral_europe -50.444      4.495  -11.223 < 2e-16 ***
Admixturegermany     -79.147      4.831  -16.384 < 2e-16 ***
Admixtureitaly_balkan_caucasus -31.819      5.259   -6.051 2.83e-09 ***
Admixturenorth_sweden -43.578      5.393   -8.080 4.89e-15 ***
Admixturerelict      -69.634      7.023   -9.915 < 2e-16 ***
Admixturesouth_sweden -48.011      4.475  -10.729 < 2e-16 ***
Admixturespain       -63.829      5.005  -12.754 < 2e-16 ***
Admixturewestern_europe -70.021      4.955  -14.132 < 2e-16 ***
FTpromotermedium     -2.274      2.426   -0.937  0.349
FTpromotershort      -6.922      5.247   -1.319  0.188
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for gaussian family taken to be 498.3158)

Null deviance: 435215  on 510  degrees of freedom
Residual deviance: 249158  on 500  degrees of freedom
AIC: 4637

Number of Fisher Scoring iterations: 2

```

Supplementary Figure 3: Longitudinal distribution of the three main *FT* promoter types.

(A, B) Test of the longitudinal distribution of the *FT* promoter types using a GLM model that includes an estimate of the population structure previously described (The 1001 Genomes Consortium 2016). The medium and long *FT* promoter types were used as a reference in (A) and (B), respectively.

Appendix

A

```
Call:
glm(formula = Lat ~ Admixture + NonCoding, data = Df)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-22.6021  -0.6762   0.2909   1.0983  14.1602

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    45.1828    0.6179   73.122 < 2e-16 ***
Admixturecentral_europe    3.0165    0.7063    4.271 2.33e-05 ***
Admixturegermany         4.2408    0.7524    5.636 2.91e-08 ***
Admixtureitaly_balkan_caucasus -3.9829    0.8171   -4.875 1.47e-06 ***
Admixturenorth_sweden     17.1817    0.8559   20.074 < 2e-16 ***
Admixturerelict          -7.4696    1.0806   -6.912 1.46e-11 ***
Admixturesouth_sweden     10.8767    0.7104   15.310 < 2e-16 ***
Admixturespain           -3.9937    0.7897   -5.057 5.98e-07 ***
Admixturewestern_europe    4.1608    0.7874    5.284 1.89e-07 ***
NonCoding-              0.9971    0.8854    1.126 0.261
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for gaussian family taken to be 12.57593)

Null deviance: 26340.3 on 510 degrees of freedom
Residual deviance: 6300.5 on 501 degrees of freedom
AIC: 2755.8

Number of Fisher Scoring iterations: 2
```

B

```
Call:
glm(formula = Long ~ Admixture + NonCoding, data = Df)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-148.692  -2.125   -0.123    6.944   123.156

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    62.5063    3.8934   16.055 < 2e-16 ***
Admixturecentral_europe   -49.9104    4.4502  -11.215 < 2e-16 ***
Admixturegermany        -78.2474    4.7411  -16.504 < 2e-16 ***
Admixtureitaly_balkan_caucasus -33.0292    5.1483   -6.416 3.25e-10 ***
Admixturenorth_sweden    -44.0242    5.3931   -8.163 2.67e-15 ***
Admixturerelict         -68.0048    6.8090   -9.987 < 2e-16 ***
Admixturesouth_sweden    -48.1695    4.4763  -10.761 < 2e-16 ***
Admixturespain          -63.9809    4.9759  -12.858 < 2e-16 ***
Admixturewestern_europe   -70.0884    4.9612  -14.127 < 2e-16 ***
NonCoding-           -0.5258    5.5788   -0.094 0.925
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for gaussian family taken to be 499.2769)

Null deviance: 435215 on 510 degrees of freedom
Residual deviance: 250138 on 501 degrees of freedom
AIC: 4637

Number of Fisher Scoring iterations: 2
```

Supplementary Figure 4: Geographical distribution of the candidate *FT* haplotypes selected for natural variation in the photoperiod control regions of *FT*.

(A, B) Test of the latitudinal and longitudinal distribution of the candidate *FT* haplotypes selected for non-coding variation, in (A) and (B), respectively, using a GLM model that includes an estimate of the population structure previously described (The 1001 Genomes Consortium 2016). The candidate *FT* haplotypes were chosen for natural variation in putative and functional TFBSs located in the photoperiod control regions of *FT*.

Position	Reference	Alteration	Accession
24331136	T	C	9552 9820
24331137	C	G	9571 9854 9862
24331157	A	AG	6911 9606 9617
24331169	A	AC	9906 9908 9911
24331278	A	T	88 403 428 430 763 765 766 772 801 1062 4807 4826 4958 5779 5822 5837 5921 5950 6025 6036 6038 6042 6046 6073 6108 6124 6154 6184 6189 6191 6192 6193 6202 6203 6252 6255 6276 6413 6445 6744 6898 6903 6929 6931 6951 6963 6969 7063 7072 7106 7126 7127 7147 7158 7169 7177 7181 7183 7186 7207 7217 7268 7305 7346 7373 7394 7413 7427 7471 7523 8227 8230 8244 8307 8357 8365 8376 8424 8426 8427 9058 9067 9070 9079 9084 9085 9089 9095 9100 9102 9105 9111 9114 9115 9125 9381 9382 9383 9394 9408 9412 9413 9416 9450 9452 9506 9531 9540 9543 9547 9548 9549 9551 9552 9554 9571 9573 9586 9599 9607 9609 9610 9611 9639 9644 9664 9665 9666 9667 9676 9677 9678 9694 9695 9696 9699 9707 9709 9716 9723 9727 9743 9745 9749 9756 9758 9759 9766 9768 9772 9774 9776 9785 9795 9798 9812 9815 9817 9820 9840 9850 9851 9854 9858 9862 9867 9869 9874 9877 9880 9890 9897 9903 9910 9912 9926 9930 9965 9966 9974 9975 9991 10009 10014 10015 14312
24331305	G	A	9711
24331307	A	G	7236
24331325	T	A	7164
24331341	A	G	9609
24331347	A	G	7323

Supplementary Table 1: Summary Table of SNP, single base insertion and deletion affecting highly conserved nucleotides at *Block A* in the corresponding accessions.

Reference and alteration correspond to the reference and alternative alleles, respectively. Alteration with two bases indicates a single base insertion. Positions are relatives to the Chromosome 1.

Appendix

Position	Reference	Alteration	Accession
24325927	A	T, G	1061 1066 1158 1166 4826 4840 6017 6042 6099 6115 6202 6284 6413 7072 7169 7268 7396 7427 8334 9057 9070 9125 9363 9412 9413 9433 9450 9452 9503 9508 9545 9553 9561 9574 9582 9689 9693 9855 9878 9955 14312 14319
24325938	A	T	9598
24325963	A	G	763 1066 1158 1166 4826 4840 6017 6042 6099 6115 6202 6284 6413 7072 7169 7217 7268 7307 7327 7396 7427 8334 9057 9070 9125 9363 9412 9413 9433 9450 9452 9503 9508 9513 9545 9553 9561 9582 9609 9687 9689 9693 9806 9809 9855 9911 9921 14312 14319
24326000	T	TA	7323
24326006	C	A	9981
24326051	A	C	2202 5741 7063 8337 9564 9714 9774 9817 9823 10018
24326055	A	T	9774
24326059	T	G	9653
24326060	G	C	9556
24326080	T	A	9879
24326106	A	G	9619 9620 9625 9626 9627 9629 9630 9631 9632 9633 9634 9636 9637 9638 9640 9641 9642 9643 9737 9951 9952 9953 9958 9959 14313 14314 14315
24326143	A	G	4840 5023 6924 9314 9791 9806 9906 9911
24326163	G	A	7218 7337 7717 9739
24326179	G	A	5779
24326211	G	A	997 6258 6974
24326220	C	G	763 765 6931 9609 9766
24326244	G	A	7520
24326245	G	C	9976
24326252	T	A	7236
24326257	C	T	6390 9130
24326258	T	C	6390 8233 9130
24326260	T	A	6390
24326265	A	C, T	4840 6924 9314 9791 9806 9908 9911
24326276	T	G	1062 6085 6113 6131 6141 6189 6203 8427 9058 9542 9569 9590 9595 9598 9607 9736 9826 9828 9835 9837 9875 9892 9897
24326357	C	A, T	801 5822 6073 6252 6255 6276 7147 7158 7208 8230 8244 8357 8365 9067 9089 9095 9100 9102 9105 9111 9114 9115 9121 9381 9382 9383 9394 9408 9416 9527 9532 9547 9548 9551 9586 9599 9663 9694 9695 9756 9757 9759 9776 9785 9815 9836 9840 9851 9858 9861 9867 9874 9877 9880 9890 9903 9975 9988 10013 10014
24326358	T	C	1062 6085 6113 6131 6141 6203 8351 8426 8427 9370 9542

Supplementary Table 2: Summary Table of SNP, single base insertion and deletion affecting highly conserved nucleotides at *Block C* in the corresponding accessions.

Reference and alteration correspond to the reference and alternative alleles, respectively. Alteration with two bases indicates a single base insertion. Bases separated by a comma indicate two alternative alleles. Positions are relative to the Chromosome 1.

Position	Reference	Alteration	Accession
24334977	A	T	5210 8240 9601 9819 980 9935
24335049	G	T	9100 9105
24335072	C	T	14312
24335140	T	A	7063
24335182	T	G	9905
24335189	A	G	6445 8419
24335195	A	AT	265 351 410 628 630 991 992 997 1002 1006 1066 1313 1552 4857 5151 5165 5210 5249 5279 5349 5395 5651 5717 5748 5757 5768 5776 5831 5836 5856 5867 6010 6011 6012 6013 6017 6019 6020 6023 6034 6040 6064 6069 6070 6071 6074 6077 6090 6091 6094 6095 6097 6098 6099 6101 6102 6104 6106 6107 6109 6111 6112 6114 6115 6118 6122 6123 6125 6126 6128 6133 6134 6137 6140 6145 6148 6149 6150 6153 6163 6166 6169 6172 6173 6174 6177 6194 6209 6210 6214 6216 6217 6218 6220 6231 6235 6238 6240 6241 6242 6244 6258 6268 6284 6296 6396 6901 6913 6918 6959 6967 6990 6997 7003 7031 7107 7111 7143 7165 7203 7288 7306 7353 7411 7424 7517 8235 8237 8240 8241 8256 8258 8259 8264 8306 8334 8335 8422 9057 9321 9332 9336 9339 9343 9363 9371 9380 9386 9388 9395 9402 9404 9405 9427 9433 9437 9442 9453 9455 9470 9471 9476 9481 9514 9518 9525 9526 9528 9529 9534 9535 9544 9546 9550 9556 9557 9558 9562 9567 9568 9576 9577 9584 9587 9589 9593 9594 9601 9602 9608 9612 9616 9672 9679 9680 9681 9682 9684 9685 9686 9687 9689 9692 9693 9732 9735 9775 9779 9799 9819 9821 9822 9824 9825 9833 9834 9836 9841 9844 9846 9847 9848 9852 9853 9856 9857 9866 9868 9874 9876 9882 9886 9891 9894 9895 9898 9899 9901 9904 9917 9918 9924 9925 9928 9933 9935 9948 10027 15560 18696
24335223	T	A	9706
24335226	T	A	9993
24335237	T	C	9979
24335239	G	C	6191 6192 6193
24335246	TG	T	9658
24335257	C	A	9558
24335273	A	T	9813
24335277	G	C	9835
24335290	C	T	7327 9513 9539 9545 9553 9555 9559 9583 9830 9843 9855 9878 9879 9888 9944
24335362	G	T	5830 6180 6434 6830 6932 7096 7209 7213 9075 9710 9741
24335369	T	G	470 476 484 504 506 531 544 546 680 681 685 687 728 742 853 854 867 868 870 915 1612 1622 1651 1652 1676 1684 1739 1741 1756 1757 1793 1797 1819 1820 1829 1834 1835 1851 1852 1853 1872 1925 1942 1943 1954 2017 2031 2053 2057 2081 2091 2106 2108 2141 2159 2166 2191 2212 2239 2240 2276 2278 2285 2286 2370 2412 5984 6024 6739 6740 6749 6750 6805 6806 6814 6927 7013 7033 7133 7160 7199 7202 7248 7250 7255 7343 7358 7359 7377 7416 7475 7515 7525 7529 7530 7566 7568 7757 7767 8037 8057 8077 8132 8233 8246 8464 8483 8699 8723 9027 9508 9536 9542 9561 9569 9578 9590 9595 9781 9826 9828 9835 9839 9845 9849 9871 9892 9915 9920 9949
24335392	T	C	9608

Supplementary Table 3: Summary Table of SNP, single base insertion and deletion affecting highly conserved nucleotides in *Block E* in the corresponding accessions.

Reference and alteration correspond to the reference and alternative alleles, respectively. Alteration with two bases indicates a single base insertion. Reference with two bases indicates a single base deletion. Bases separated by a comma announce two alternative alleles. Positions are relatives to the Chromosome 1.

Appendix

Purpose	ID	Name	Sequence	Orientation	Comments
Pyrosequencing	T052	Pyr_R1_bio	(Bio)AG CCACTCTCCCTCTGACAATT	R	Biotinylated
	T053	FII_gDNA	GACTGATATCCCTGCTACAACTGG	F	
	T054	Pyr_S1	GTTTCGACAGCTTGG	F	Sequencing primer
Cloning	T854	BsaI-FTislandfw	GGGTCTCT AGCAAAAACCTTGTAACGCTAG	F	Promoter variant into <i>pDONR207</i>
	T855	BsaI-FTislandrv	GGGTCTCT TGCTAAATCTCATTATGTGA	R	
	T856	BsaI5.7kbFTfw	GGGTCTCC ATTGTGTAACAAAAATCT	F	
	T857	BsaI-Ftprom-re	GGGTCTCCT TTGATCTTGAACAAACAGGT	R	
	T858	pGG-ENTRY-fw	GGGTCTCT CAAAGTACCCAGCTTCTTGTACAA	F	
	T859	pGG-ENTRY-re	GGGTCTCC AAATGTAAGCCTGCTTTTGTAGCA	R	
	T024	OLE_Gre_F	GTGCTGCAAGGCGATTAAGTTGGG	F	<i>OLE1</i> into <i>pGreen</i>
	T025	OLE_Gre_R	ATCCCCCTTTCGCCAGCTGG	R	
	T026	O-Gre1_F	CCAGCTGGCGAAAGGGGGAT GGGCCCTACTTAGATCAAC	F	
	T027	O-Gre1_R	CCAACTTAATCGCCTTGCAGCA CAGGTCACTGGATTTGGTTTATAG	R	
	T083	ale_cat_F	CCATAAAGGATTGGATGAGTGC	F	CCAAT-box point mutation
	T084	ale_cat_R	GCACTCATCCAATCCTTATGG	R	
	T087	FaL_S3_F	GGTGGGTTTGATACCAC	F	Shadow3A point mutation
	T088	FaL_S3_R	GTGGTATTCCAAACCCACC	R	
RT	T055	Oligo(dT)	TTTTTTTTTTTTTTTTT		
Gene Expression	T050	PP2A_re_RT	AAATACGCCCAACGAACAAA	R	<i>PP2A</i>
	T051	PP2A_fw_RT	CAGCAACGAATTGTGTTGG	F	
	J116	RT-FT-cDNA	GGTGGAGAAGACCTCAGGAA	F	<i>FT</i>
	J117	RT-FT-cDNA	ACCCTGGTGCAATACACTGTT	R	
TE insertion genotyping	F162	3prime	TGGCCGCAGTTTCTACAAT	F	Downstream <i>FT</i> 3'UTR
	J166	Vdr	AGAAAATCAAAGGGGTTTCGAG	R	
	Z215	DownstreamBlockCF2	AACAAGCACTTATGTGTGTTAAGTCT	F	
	Z216	DownstreamBlockCR2	CCCAAATCTCGTTGCCTAAA	R	Downstream of <i>Block C</i>
	Z203	Swedish_GenoF1	AGACGTCTCTACTTGGTTGCG	F	
	Z204	Swedish_GenoR1	TGCCCTGAGAAAACCGTAAGA	R	
	Z208	UpstreamBlockCF1	TCGTCACTGCTTCTGACGTTT	F	Upstream of <i>Block C</i>
	Z209	UpstreamBlockCR1	CCGATTGAATTGTTGTTGGT	R	
Tandem Repeat Insertion	KK162	AT1G13320/PP2A	ACACAATTCGTTGCTGCTTCT	F	<i>PP2A</i>
	KK163	AT1G13320/PP2A	TGCTTGGTGGAGCTAAGTGA	R	
	KK164	PPT	CCAGTCCCGTGCTTGAA	F	<i>PPT</i>
	KK165	PPT	AGAGCGTGGTCGCTGTCAT	R	

Supplementary Table 4: List of primers used in this study.

ID numbers correspond to the stock primers from Dr. Franziska Turck's lab. The primers T854, T855, T856 and T857 were kindly provided by Dr. Franziska Turck. The primers KK162, KK163, KK164 and KK165 were kindly provided by Kristin Krause. Nucleotides in bold and bold italic correspond to the BsaI recognition sites and the overlapping sequences required for the Golden Gate Assembly and the Gibson Assembly® reaction, respectively.

Supplementary Table 5: Summary Table of the natural variations identified at *FT* and *FRI* (pages 95 to 116).

Lat and Long correspond to the latitude and longitude, respectively. + represents the presence of the Col-0 and Sf-2 allele of *FT* and *FRI*, respectively. NA is indicated when the characterization was not possible. Accessions are indicated with their ID number.

Accession	Country	Lat	Long	FT promoter	Coding SNP FT	Noncoding SNP FT	TEP FT	FRI
88	FRA	47,4	0,683333	NA	+	+	+	NA
108	FRA	48,5167	-4,06667	NA	+	+	+	-
139	FRA	48,5167	-4,06667	NA	+	+	+	-
159	FRA	47,35	3,93333	long	+	+	+	-
350	FRA	46,6667	4,11667	NA	+	+	+	-
351	FRA	46,6667	4,11667	NA	+	+	+	+
403	CZE	49,3667	16,2667	NA	+	+	+	-
410	CZE	49,4211	16,3497	NA	+	+	+	+
424	CZE	49,4112	16,2815	long	+	+	+	+
428	CZE	49,403	16,232	NA	+	+	+	+
430	AUT	47	15,5	medium	+	+	+	-
630	USA	40,7777	-72,9069	NA	+	-	+	+
763	KGZ	42,3	74,3667	medium	+	+	+	-
765	KGZ	42,1833	73,4	medium	+	+	+	+
766	KGZ	42,5833	73,6333	medium	+	+	+	+
768	KGZ	42,8	76,35	medium	+	+	+	+
770	KGZ	40,046526	72,683613	NA	+	+	+	NA
772	TJK	37,35	72,4667	medium	+	+	+	-
801	USA	37,9169	-84,4639	NA	+	+	+	+
870	USA	41,8266	-86,4366	medium	+	+	+	+
915	USA	41,8972	-71,4378	NA	+	+	+	-
932	USA	42,3634	-71,1445	NA	+	+	+	+
991	SWE	55,3833	14,05	medium	+	+	+	+
992	SWE	55,3833	14,05	medium	+	+	+	+
997	SWE	55,3833	14,05	medium	-	+	+	+
1002	SWE	55,3833	14,05	medium	+	+	+	+
1061	SWE	55,7167	14,1333	NA	+	+	+	+
1062	SWE	55,7167	14,1333	medium	+	+	+	+
1063	SWE	55,7167	14,1333	medium	+	+	+	+
1066	SWE	55,7167	14,1333	NA	+	+	+	+
1074	SWE	55,7167	14,1333	medium	+	+	+	+
1137	SWE	56,6167	16,65	NA	+	+	+	+
1158	SWE	56,7	16,5167	medium	+	+	+	+
1254	SWE	59,4333	17,0167	medium	+	+	+	+
1257	SWE	59,4333	17,0167	medium	+	+	+	+
1318	SWE	59,5667	16,8667	NA	+	+	+	NA
1363	SWE	59,7833	17,5833	NA	+	+	+	NA
1367	SWE	59,7833	17,5833	NA	+	+	+	NA
1435	SWE	62,8	18,2	medium	-	+	+	-
1552	SWE	63,0833	18,3667	medium	+	+	+	+
1585	SWE	65,25	15,6	NA	+	+	+	NA
1829	USA	42,051	-86,509	NA	+	+	+	+
1853	USA	43,595	-86,2657	NA	+	+	+	+
1872	USA	43,595	-86,2657	medium	+	+	+	+
1890	USA	43,5139	-86,1859	long	+	+	+	+
1925	USA	43,5251	-86,1843	medium	+	+	+	-
1954	USA	43,5187	-86,1739	NA	+	+	+	-

Appendix

Accession	Country	Lat	Long	FT promoter	Coding SNP FT	Noncoding SNP FT	TEP FT	FRI
2171	USA	42,148	-86,431	NA	+	+	+	+
2278	USA	43,665	-86,496	medium	+	+	+	+
4779	UK	50,4	-4,9	medium	+	+	+	+
4807	UK	50,4	-4,9	NA	+	+	+	+
4826	UK	50,4	-4,9	NA	+	+	+	-
4884	UK	50,3	-4,9	NA	+	+	+	-
4900	UK	50,3	-4,8	NA	+	+	+	-
4931	UK	50,327643	-4,6	NA	+	+	+	NA
4958	UK	50,5	-4,5	NA	+	+	+	+
5023	UK	51,3	0,5	long	-	+	+	-
5104	UK	51,3	0,5	long	+	+	+	-
5210	UK	51,2	0,3	medium	+	+	+	+
5236	UK	51,2	0,4	NA	+	+	+	+
5253	UK	51,1	0,6	NA	+	+	+	-
5276	UK	51,3	1,1	NA	+	+	+	-
5349	UK	51,1	0,4	NA	+	-	+	NA
5353	UK	54,5	-3	medium	+	+	+	+
5470	UK	54,7	-3,4	NA	+	+	+	NA
5486	UK	54,6	-3,3	long	+	+	+	-
5506	UK	54,6	-3,1	NA	+	+	+	NA
5535	UK	54,6	-3,1	NA	+	+	+	NA
5577	UK	54,7	-3,4	NA	+	+	+	+
5644	UK	54,4	-2,9	NA	+	+	+	+
5726	UK	51,3	1,1	NA	+	+	+	+
5741	UK	56,6	-4,1	NA	+	+	+	-
5748	UK	56	-4,4	medium	+	+	+	-
5778	UK	52,2	1,5	NA	+	+	+	NA
5779	UK	51	-3,1	NA	-	+	+	-
5800	UK	57,4	-5,5	long	+	+	+	-
5818	UK	51,8	-0,6	NA	+	+	+	NA
5829	SWE	55,38383	14,06125	NA	+	+	+	NA
5830	SWE	56,3333	15,9667	long	+	+	+	+
5831	SWE	56,3333	15,9667	medium	+	+	+	+
5832	SWE	56,3333	15,9667	long	+	+	-	-
5835	SWE	63,324	18,484	NA	+	+	+	NA
5837	CZE	49,4013	16,2326	medium	+	+	+	+
5856	SWE	63,0167	17,4914	NA	+	+	+	+
5865	SWE	55,76	14,12	medium	+	+	+	+
5867	SWE	55,76	14,12	medium	+	+	+	+
5874	CZE	49,4112	16,2815	NA	+	+	+	+
5890	CZE	49,4112	16,2815	NA	+	+	+	-
5893	CZE	49,4112	16,2815	medium	+	+	+	-
5907	CZE	49,4112	16,2815	NA	+	+	+	+
5921	CZE	49,4112	16,2815	NA	+	+	+	+
5950	CZE	49,4112	16,2815	NA	+	+	+	+
5964	CZE	49,4112	16,2815	NA	+	+	+	NA
5984	CZE	49,4112	16,2815	NA	+	+	+	+

Accession	Country	Lat	Long	FT promoter	Coding SNP FT	Noncoding SNP FT	TEP FT	FRI
5993	CZE	49,4112	16,2815	NA	+	+	+	+
6008	CZE	49,1	16,2	medium	+	+	+	-
6009	SWE	62,877	18,177	NA	+	+	+	+
6011	SWE	62,877	18,177	medium	+	+	+	+
6012	SWE	62,877	18,177	medium	+	+	+	+
6013	SWE	62,877	18,177	NA	+	+	+	+
6016	SWE	62,9	18,4	NA	+	+	+	+
6017	SWE	62,9	18,4	medium	+	+	+	+
6019	SWE	56,06	14,29	medium	+	+	+	+
6020	SWE	56,06	14,29	medium	+	+	+	+
6021	SWE	56,06	14,29	short	+	+	+	+
6022	SWE	56,06	14,29	NA	+	+	-	NA
6023	SWE	55,7509	13,3712	medium	+	+	+	+
6024	SWE	55,7509	13,3712	medium	+	+	+	+
6025	SWE	62,6437	17,7339	medium	+	+	+	+
6030	SWE	62,806	18,1896	NA	+	-	+	+
6034	SWE	56,1	13,74	NA	+	+	+	+
6035	SWE	56,1	13,74	NA	+	+	+	+
6036	SWE	56,1	13,74	NA	+	+	+	+
6038	SWE	56,1	13,74	medium	+	+	+	+
6040	SWE	55,66	13,4	NA	+	+	+	+
6041	SWE	56,0328	14,775	long	+	+	+	+
6043	SWE	62,801	18,079	NA	+	+	+	+
6046	SWE	62,801	18,079	NA	+	+	+	+
6064	SWE	62,9513	18,2763	NA	+	+	+	+
6069	SWE	62,9513	18,2763	medium	+	+	+	+
6070	SWE	62,9308	18,3448	medium	+	+	+	+
6071	SWE	62,9308	18,3448	medium	+	+	+	+
6073	SWE	56,1481	15,8155	medium	+	+	+	+
6074	SWE	56,4573	16,1308	NA	+	+	+	+
6076	SWE	55,6942	13,4504	NA	+	+	+	+
6085	SWE	55,7097	13,2145	medium	+	+	-	-
6090	SWE	55,6525	13,2197	medium	+	+	+	+
6092	SWE	55,6514	13,2233	NA	+	+	+	+
6094	SWE	55,6494	13,2147	medium	+	+	+	+
6097	SWE	55,6481	13,2264	medium	+	+	+	+
6098	SWE	55,6561	13,2178	medium	+	+	+	+
6099	SWE	55,6575	13,2386	NA	+	+	+	+
6100	SWE	55,6	13,2	medium	+	+	+	+
6102	SWE	55,6	13,2	medium	+	+	+	+
6104	SWE	55,7	13,2	medium	+	+	+	+
6105	SWE	55,7967	13,1211	NA	+	+	+	-
6106	SWE	55,7931	13,1186	medium	+	+	+	+
6107	SWE	55,7942	13,1222	medium	+	+	+	+
6108	SWE	55,7989	13,1206	medium	+	+	+	-
6109	SWE	55,7936	13,1233	medium	+	+	+	+
6111	SWE	55,7989	13,1219	medium	+	+	+	+

Appendix

Accession	Country	Lat	Long	FT promoter	Coding SNP <i>FT</i>	Noncoding SNP <i>FT</i>	TEP <i>FT</i>	<i>FRI</i>
6112	SWE	55,7967	13,1044	NA	+	+	+	+
6113	SWE	55,8078	13,1028	NA	+	+	+	+
6114	SWE	55,8097	13,1342	medium	+	+	+	+
6115	SWE	55,8	13,1367	medium	+	+	-	+
6118	SWE	55,7	13,2	medium	+	+	+	+
6122	SWE	55,8364	13,3075	medium	+	+	+	+
6124	SWE	55,8378	13,3092	medium	+	+	+	+
6125	SWE	55,8403	13,3106	NA	+	+	+	+
6126	SWE	55,8411	13,3047	medium	+	+	+	+
6127	SWE	55,8428	13,3058	NA	+	+	+	-
6128	SWE	55,8397	13,2881	medium	+	+	+	+
6129	SWE	55,8544	13,2875	NA	+	+	+	NA
6131	SWE	55,8369	13,3181	medium	+	+	-	+
6132	SWE	55,8386	13,3186	NA	+	+	+	+
6133	SWE	55,8364	13,2906	medium	+	+	+	+
6134	SWE	55,8383	13,2906	medium	+	+	+	+
6136	SWE	55,9336	13,5519	NA	+	+	+	+
6137	SWE	55,9419	13,5603	NA	+	+	+	+
6138	SWE	55,9403	13,5511	medium	+	+	+	+
6140	SWE	55,9392	13,5539	medium	+	+	+	+
6141	SWE	55,9414	13,5542	NA	+	+	-	-
6142	SWE	55,9428	13,5558	medium	+	+	+	+
6145	SWE	55,9497	13,5533	medium	+	+	+	+
6148	SWE	55,9319	13,5508	NA	+	+	+	+
6149	SWE	55,9281	13,5481	NA	+	+	+	+
6150	SWE	55,9261	13,5319	medium	+	+	+	+
6151	SWE	55,6528	13,2244	medium	+	+	+	+
6153	SWE	62,6425	17,7422	NA	+	+	+	+
6154	SWE	62,6422	17,7406	NA	+	+	+	+
6163	SWE	62,6425	17,7356	medium	+	+	+	+
6166	SWE	62,6425	17,7372	medium	+	+	+	+
6169	SWE	62,8714	18,3447	NA	+	+	+	+
6171	SWE	62,8717	18,3444	NA	+	+	+	+
6172	SWE	62,8717	18,3436	NA	+	-	+	+
6173	SWE	62,8717	18,3419	medium	+	-	+	+
6174	SWE	62,8719	18,3422	medium	+	-	+	+
6177	SWE	62,6322	17,69	medium	+	+	+	+
6180	SWE	62,6322	17,6906	NA	+	+	+	-
6184	SWE	62,8892	18,4522	medium	+	+	+	+
6188	SWE	55,7683	14,1386	long	+	-	+	+
6189	SWE	55,7686	14,1383	NA	+	+	+	+
6191	SWE	55,7689	14,1375	medium	+	+	+	-
6192	SWE	55,7692	14,1369	NA	+	+	+	+
6193	SWE	55,7694	14,1347	NA	+	+	+	-
6194	SWE	55,7706	14,1342	medium	+	+	+	-
6195	SWE	55,7708	14,1342	long	+	+	+	+
6197	SWE	55,7714	14,1333	NA	+	+	+	NA

Accession	Country	Lat	Long	FT promoter	Coding SNP FT	Noncoding SNP FT	TEP FT	FRI
6198	SWE	55,7708	14,1331	long	+	+	+	+
6201	SWE	55,7719	14,1211	short	+	+	+	+
6202	SWE	55,7717	14,1206	NA	+	+	+	+
6203	SWE	55,7714	14,1208	medium	+	+	+	+
6209	SWE	62,8836	18,1842	medium	+	+	+	+
6210	SWE	62,8839	18,1836	NA	+	+	+	-
6216	SWE	63,0167	18,3283	medium	-	+	+	-
6217	SWE	63,0169	18,3283	medium	-	+	+	+
6218	SWE	63,0172	18,3283	medium	+	+	+	+
6220	SWE	62,806	18,1896	medium	+	-	+	+
6221	SWE	62,806	18,1896	medium	+	+	+	+
6231	SWE	62,96	18,2844	NA	+	+	+	+
6235	SWE	62,9611	18,3589	medium	+	+	+	+
6237	SWE	62,9619	18,35	NA	+	+	+	+
6238	SWE	62,9619	18,35	NA	+	+	+	+
6240	SWE	62,9622	18,35	NA	+	+	+	+
6241	SWE	62,9614	18,3608	NA	+	+	+	+
6242	SWE	55,7	13,2	medium	+	+	+	+
6243	SWE	56,27373	13,90045	medium	+	+	+	+
6244	SWE	62,9169	18,4728	NA	+	+	+	+
6252	SWE	55,5796	14,3336	medium	+	+	+	+
6255	SWE	55,5796	14,334	NA	+	+	+	+
6268	SWE	55,5796	14,3336	medium	-	+	+	+
6276	SWE	55,5796	14,3336	medium	+	+	+	+
6284	SWE	55,5796	14,3336	medium	+	+	+	+
6288	CZE	49,2771	16,6314	NA	+	+	+	NA
6296	CZE	49,2771	16,6314	NA	+	+	+	+
6390	CZE	49,2771	16,6314	NA	+	+	+	-
6396	CZE	49,2771	16,6314	NA	+	+	+	+
6413	SWE	56,06	13,97	medium	+	+	+	+
6424	CZE	49,3853	16,2544	NA	+	+	+	+
6445	CZE	49,3853	16,2544	NA	+	+	+	-
6680	GER	51,85	6,4333	medium	+	+	+	+
6709	USA	47,6479	-122,305	NA	+	+	+	NA
6744	USA	40,8585	-73,4675	NA	+	+	+	+
6897	FRA	45	1,3	medium	+	+	+	+
6900	SWE	63,324	18,484	NA	+	-	+	+
6901	SWE	63,324	18,484	NA	+	+	+	+
6903	CZE	49,4013	16,2326	medium	+	+	+	+
6904	CZE	49,2	16,6166	short	+	+	+	+
6906	POR	40,2077	-8,42639	NA	+	+	+	NA
6908	UK	51,4083	-0,6383	medium	+	+	+	-
6909	USA	38,3	-92,3	long	+	+	+	-
6911	CPV	15,1111	-23,6167	medium	+	+	+	+
6913	SWE	62,877	18,177	NA	+	+	+	+
6915	GER	50,3	6,3	long	+	+	+	+
6916	RUS	58,3	25,3	NA	+	+	+	+

Appendix

Accession	Country	Lat	Long	FT promoter	Coding SNP <i>FT</i>	Noncoding SNP <i>FT</i>	TEP <i>FT</i>	<i>FRI</i>
6917	SWE	63,0165	18,3174	NA	+	+	+	+
6918	SWE	63,0165	18,3174	medium	+	+	+	+
6919	GER	50,3	8	medium	+	+	+	-
6921	GER	51,5338	9,9355	NA	+	+	+	NA
6922	GER	50,3	8	long	+	+	+	-
6923	UK	51,4083	-0,6383	medium	+	+	+	-
6924	UK	51,4083	-0,6383	long	-	+	+	-
6926	USA	44,46	-85,37	medium	+	+	+	-
6927	USA	41,2816	-86,621	medium	+	+	+	+
6928	USA	41,2816	-86,621	NA	+	+	+	NA
6929	TJK	38,48	68,49	medium	+	+	+	+
6931	KAZ	7518	73,1	NA	+	+	+	+
6933	ESP	41,59	2,49	long	+	+	+	+
6936	FRA	46	3,3	NA	+	+	+	NA
6937	GER	49	9,3	NA	+	+	+	NA
6938	RUS	55,7522	37,6322	long	+	+	+	+
6939	LIB	32,34	22,46	NA	+	+	+	NA
6940	GER	50,3	8,3	long	+	+	+	-
6943	UK	51,4083	-0,6383	medium	+	+	+	+
6944	UK	51,4083	-0,6383	medium	+	+	+	-
6945	NED	52,24	4,45	long	+	+	+	+
6951	CZE	49,42	16,36	medium	+	+	+	+
6956	CZE	49,42	16,36	NA	+	+	+	+
6957	CZE	49,42	16,36	NA	+	+	+	+
6958	FRA	46	3,3	medium	+	+	+	-
6959	FRA	48,5	-1,41	medium	+	+	+	-
6960	FRA	48,5	-1,41	NA	+	+	+	-
6961	ESP	38,3333	-3,53333	NA	+	+	+	+
6963	TJK	38,35	68,48	medium	+	+	+	+
6964	SWE	56,31635	16,03529	NA	+	+	+	+
6965	SWE	56,31635	16,03529	NA	+	+	+	+
6966	UK	51,4083	-0,6383	medium	+	-	+	+
6967	UK	51,4083	-0,6383	medium	+	+	+	-
6968	FIN	60	23,5	medium	+	+	+	+
6969	FIN	60	23,5	medium	+	+	+	+
6970	ESP	41,7194	2,93056	long	+	+	+	+
6971	ESP	41,7194	2,93056	medium	+	+	+	+
6972	JPN	34,43	136,31	NA	+	+	+	NA
6973	SWE	56,0648	13,9707	long	+	+	+	+
6974	SWE	56,0648	13,9707	NA	-	+	+	+
6975	AUT	48,3	14,45	medium	+	+	+	-
6976	AUT	48,3	14,45	NA	+	+	+	+
6979	SUI	47,25	8,26	medium	+	+	+	-
6982	GER	52,3	9,3	NA	+	+	+	-
6984	CZE	49,3853	16,2544	medium	+	+	+	-
6985	CZE	49,3853	16,2544	NA	+	+	+	NA
6986	UK	57,1539	-2,2207	medium	+	+	+	+

Accession	Country	Lat	Long	FT promoter	Coding SNP FT	Noncoding SNP FT	TEP FT	FRI
6989	UK	54,8	-2,4333	medium	+	+	+	+
6994	FRA	45,9	6,13028	NA	+	+	+	NA
6997	NED	51,8333	5,5833	long	+	+	+	-
7000	GER	50,9167	9,57073	medium	+	+	+	-
7002	NED	51,3333	6,1	long	+	+	+	+
7008	NED	52	5,675	long	+	+	+	+
7010	GER	49,6803	8,6161	NA	+	+	+	NA
7013	GER	52,4584	13,287	medium	+	+	+	+
7014	UK	56,5459	-4,79821	long	+	+	+	+
7025	ITA	44,5041	11,3396	medium	+	+	+	+
7026	UK	54,4	-3,2667	medium	+	+	+	-
7031	GER	50,0167	8,6667	medium	+	-	+	-
7033	USA	41,3599	-122,755	NA	+	+	+	-
7036	GER	50,5	9,5	long	+	+	+	-
7036	SWE	56,4666	16,1284	long	+	+	+	+
7062	GER	50,2981	8,26607	long	+	+	+	-
7063	ESP	29,2144	-13,4811	NA	+	+	+	+
7064	UK	51,3	1,1	medium	+	+	+	+
7067	ITA	37,3	15	NA	+	+	+	-
7068	ITA	42	12,1	medium	+	+	+	+
7071	FRA	48,0717	1,33867	long	+	+	+	-
7075	FRA	43,3779	2,54038	NA	+	+	+	NA
7077	POR	40,12	-8,25	short	+	+	+	+
7081	POR	40,2077	-8,42639	medium	+	+	+	+
7092	FRA	49,416	2,823	medium	+	+	+	+
7094	GER	49,8724	8,65081	long	+	+	+	-
7098	FRA	47	5	NA	+	+	+	NA
7102	GER	50,7224	8,2372	medium	+	+	+	+
7103	CZE	49,4167	16,2667	medium	+	+	+	+
7106	GER	51,051	13,7336	medium	+	+	+	-
7107	UK	54,7761	-1,5733	medium	+	+	+	+
7109	UK	51,3	0,5	medium	+	+	+	+
7117	GER	51,5105	9,68253	NA	+	+	+	-
7119	GER	50	8,5	long	+	+	+	-
7120	GER	50	8,5	medium	+	+	+	-
7123	GER	50,1721	8,38912	NA	+	+	+	NA
7125	GER	49,5955	11,0087	long	+	+	+	-
7130	FRA	44,6447	2,56481	short	+	+	+	+
7133	GER	50,1102	8,6822	medium	+	+	+	-
7143	NED	51,0167	5,86667	medium	+	+	+	-
7148	JPN	35,45	137,42	NA	+	+	+	NA
7158	AUT	47	15,5	NA	+	+	+	+
7160	USA	43,178	-85,2532	medium	+	+	+	+
7161	GER	53,5	10,5	long	+	+	+	-
7162	GER	52,24	9,44	long	+	+	+	-
7164	DEN	55,675	12,5686	long	+	+	+	+
7165	GER	51,3472	8,28844	medium	+	+	+	-

Appendix

Accession	Country	Lat	Long	FT promoter	Coding SNP <i>FT</i>	Noncoding SNP <i>FT</i>	TEP <i>FT</i>	<i>FRI</i>
7166	NED	51,25	5,9	NA	+	+	+	NA
7177	CZE	49	15	medium	+	+	+	-
7186	LTU	54,8969	23,8924	NA	+	+	+	+
7192	UK	55,6395	-5,66364	medium	+	+	+	-
7199	GER	50,95	6,9666	medium	+	+	+	-
7202	GER	50,1797	8,50861	medium	+	+	+	-
7203	GER	49,631	11,5722	medium	+	+	+	-
7206	GER	50,0742	8,96617	NA	+	+	+	NA
7207	JPN	35,0085	135,752	medium	+	+	+	+
7208	UK	55,6739	-3,78181	NA	+	+	+	-
7217	FRA	48	0,5	medium	+	+	+	-
7223	GER	50,3833	8,0666	medium	-	+	+	+
7231	GER	50,3833	8,0666	NA	+	+	+	-
7236	LTU	NA	NA	medium	+	+	+	+
7244	GER	50,001	8,26664	medium	+	+	+	-
7248	USA	41,3923	-70,6652	NA	+	+	+	-
7250	GER	51,9183	10,1138	NA	+	+	+	-
7252	UK	54,6167	-2,3	NA	+	+	+	NA
7255	POL	50,95	20,5	medium	+	+	+	-
7258	GER	50,5	8,5	long	+	+	+	-
7263	NZL	-37,7871	175,283	NA	+	+	+	NA
7273	GER	51,0581	13,2995	NA	+	+	+	-
7276	GER	50,2	8,5833	long	+	+	+	-
7280	GER	53,1667	8,2	long	+	+	+	-
7282	GER	50,3827	8,01161	medium	+	+	+	+
7287	GER	53,3422	8,42255	long	+	+	+	-
7296	RUS	59	29	long	+	+	+	-
7298	AUT	47,04	10,51	medium	+	+	+	-
7306	CAN	49,2655	-123,206	medium	+	+	+	+
7307	FRA	48,0653	-2,96591	NA	+	+	+	-
7308	GER	50,7167	7,1	NA	+	+	+	NA
7320	FRA	49,4424	1,09849	medium	+	-	+	+
7322	RUS	56,3	34	long	+	+	+	-
7323	UKR	49	38,28	medium	-	+	+	+
7327	ESP	41,7833	3,03333	long	+	+	+	+
7328	ESP	41,7833	3,03333	NA	+	+	+	+
7329	USA	37,21	-121,16	NA	+	+	+	NA
7330	JPN	43,0553	141,346	NA	+	+	+	NA
7332	USA	47	-122,2	medium	+	+	+	-
7333	ITA	46,5438	11,5614	long	+	+	+	-
7337	GER	50,8738	8,02341	long	-	+	+	-
7343	GER	52,5339	13,181	medium	+	+	+	+
7346	GER	52,6058	11,8558	NA	+	+	+	+
7347	RUS	52	36	medium	+	+	+	+
7349	CZE	49,5	14,5	long	+	+	+	+
7350	USA	47,2413	-122,459	long	+	+	+	+
7351	UK	56,4278	-5,23439	NA	+	+	+	NA

Accession	Country	Lat	Long	FT promoter	Coding SNP FT	Noncoding SNP FT	TEP FT	FRI
7354	SWE	56,5	14,9	medium	+	+	+	+
7355	ITA	41,96	12,8	NA	+	+	+	NA
7356	USA	41,6639	-83,5553	long	+	+	+	+
7372	AUT	47,0748	9,9042	long	+	+	+	-
7377	USA	43,2708	-85,2563	medium	+	+	+	+
7383	CAN	49,2655	-123,206	medium	+	+	+	+
7387	UK	54,9902	-2,3671	medium	+	+	+	-
7396	RUS	52,3	30	medium	+	+	+	+
7411	GER	47,9299	10,8134	medium	+	-	+	-
7413	LTU	54,6833	25,3167	NA	+	+	+	+
7415	GER	49,7878	9,9361	long	+	+	+	+
7416	USA	37,45	-119,35	medium	+	+	+	+
7417	SUI	47,3667	8,55	NA	+	+	+	+
7419	GER	50,3058	8,32213	long	+	+	+	-
7424	CZE	49,2	16,6166	medium	+	+	+	-
7427	DEN	NA	NA	medium	+	+	+	-
7438	RUS	61,36	34,15	NA	+	+	+	NA
7458	DEN	55,675	12,5687	NA	+	+	+	NA
7460	CZE	NA	NA	medium	+	+	+	+
7461	CZE	49	15	long	+	+	+	-
7515	USA	41,5609	-86,4251	medium	+	+	+	+
7516	SWE	55,58	14,334	medium	+	+	+	+
7517	SWE	55,58	14,334	NA	+	+	+	+
7518	SWE	56,14	15,78	NA	-	+	+	+
7519	SWE	56,14	15,78	medium	+	+	+	+
7520	CZE	49,38	16,81	medium	-	+	+	+
7521	CZE	49,38	16,81	medium	+	+	+	+
7522	ITA	44,15	9,65	NA	+	+	+	NA
7523	USA	42,0945	-86,3253	NA	+	+	+	-
7524	USA	42,036	-86,511	NA	+	+	+	NA
7525	USA	42,036	-86,511	medium	+	+	+	+
7526	USA	42,0945	-86,3253	NA	+	+	+	NA
7717	USA	41,273	-86,625	long	-	+	+	+
7917	USA	42,0945	-86,3253	long	+	+	+	+
7947	USA	42,0945	-86,3253	NA	+	+	+	+
8077	USA	41,3423	-86,7368	medium	+	+	+	-
8132	USA	42,036	-86,511	NA	+	+	+	+
8213	ESP	43,25	-6	NA	+	+	+	NA
8214	FRA	49	2	medium	+	+	+	+
8222	SWE	56,0328	14,775	medium	+	+	+	-
8227	SWE	62,7989	17,9103	medium	+	+	+	+
8230	SWE	56,68	16,5	NA	+	+	+	+
8231	SWE	56,3	16	NA	+	+	+	+
8233	USA	41,1876	-87,1923	medium	+	+	+	+
8234	SWE	56,4606	15,8127	long	+	+	+	+
8235	CZE	48,8	17,1	medium	+	+	+	+
8236	CZE	49,33	15,76	medium	+	+	+	+

Appendix

Accession	Country	Lat	Long	FT promoter	Coding SNP <i>FT</i>	Noncoding SNP <i>FT</i>	TEP <i>FT</i>	<i>FRI</i>
8237	SWE	55,8	13,1	NA	+	+	+	+
8238	UK	51,15	0,4	long	+	+	+	-
8239	GER	51	7	long	+	+	+	-
8240	SWE	55,705	13,196	NA	+	+	+	-
8241	SWE	55,9473	13,821	medium	+	+	+	+
8242	SWE	56,1494	15,7884	medium	+	+	+	+
8243	ITA	43,7703	11,2547	medium	+	+	+	+
8244	FRA	48,6103	2,3086	medium	+	+	+	+
8246	USA	35	-79,18	medium	+	+	+	+
8247	SWE	56,07	13,74	long	+	+	+	+
8249	SWE	57,7	15,8	long	+	+	+	+
8258	SWE	56,4	12,9	medium	+	+	+	+
8259	SWE	56,4	12,9	medium	+	+	+	+
8264	ESP	41,6833	2,8	medium	+	+	+	+
8266	SWE	55,86	13,51	NA	+	+	+	NA
8275	FRA	49	0,5	NA	+	+	+	NA
8283	SWE	55,76	14,12	long	+	+	+	+
8284	CZE	49,4112	16,2815	NA	+	+	+	+
8285	CZE	49,4112	16,2815	NA	+	+	+	-
8290	GER	50	8,5	medium	+	+	+	-
8297	SUI	46,5	6,08	long	+	+	+	+
8306	SWE	56,1	13,74	medium	+	+	+	+
8307	SWE	56,1	13,74	medium	+	+	+	+
8311	AUT	47,5	11,5	medium	+	+	+	+
8312	GER	50,5	7,5	long	+	+	+	-
8325	POL	50	19,3	NA	+	+	+	NA
8326	SWE	56,0328	14,775	medium	+	+	+	+
8335	SWE	55,71	13,2	medium	+	+	+	+
8337	ITA	44	12,37	NA	+	+	+	+
8351	SWE	60,25	18,37	medium	+	+	+	+
8353	ITA	38,07	13,22	NA	+	+	+	NA
8354	RUS	58	56,3167	long	+	+	+	+
8357	ESP	41,5	2,25	medium	+	+	+	+
8365	CZE	49	16	medium	+	+	+	-
8369	SWE	55,6942	13,4504	medium	+	+	+	+
8376	SWE	62,69	18	medium	+	+	+	+
8378	CZE	49,49	14,24	NA	+	+	+	NA
8386	SWE	58,9	11,2	long	+	+	+	+
8387	SWE	59	18	long	+	+	+	+
8419	LTU	54,6833	25,3167	medium	+	+	+	+
8420	GER	50,0667	8,5333	long	+	+	+	-
8422	SWE	56,06	14,29	medium	+	+	+	-
8423	SWE	56,1	13,74	NA	+	+	+	NA
8424	IND	35	77	medium	+	+	+	+
8424	IND	35	77	medium	+	+	+	+
8424	IND	35	77	medium	+	+	+	+
8424	IND	35	77	medium	+	+	+	+

Accession	Country	Lat	Long	FT promoter	Coding SNP FT	Noncoding SNP FT	TEP FT	FRI
8426	SWE	56,06	13,97	medium	+	+	+	-
8428	AUT	48,3	14,45	NA	+	+	+	NA
8430	NED	52,25	4,5667	NA	+	+	+	NA
8472	USA	41,6862	-86,8513	NA	+	+	+	NA
8584	USA	42,093	-86,359	NA	+	+	+	NA
9045	USA	42,039	-86,5154	NA	+	+	+	NA
9057	SWE	56,1	13,9167	medium	+	+	+	-
9061	TUR	38,3011	42,2239	NA	+	+	+	NA
9063	TUR	38,3011	42,2239	NA	+	+	+	NA
9064	TUR	38,3011	42,2239	NA	+	+	+	NA
9066	AZE	38,6536	48,7992	NA	+	+	+	NA
9067	AZE	38,6536	48,7992	medium	+	+	+	+
9069	AZE	38,6536	48,7992	medium	+	+	+	+
9070	AZE	38,6536	48,7992	NA	+	+	+	+
9073	AZE	38,7406	48,6131	NA	+	+	+	NA
9075	AZE	38,7406	48,6131	long	+	+	+	+
9078	AZE	38,7406	48,6131	medium	+	+	+	+
9079	AZE	38,7833	48,5517	NA	+	+	+	+
9081	AZE	38,7833	48,5517	NA	+	+	+	+
9082	AZE	38,7833	48,5517	NA	+	+	+	NA
9084	AZE	38,7833	48,5517	NA	+	+	+	+
9085	AZE	38,7833	48,5517	NA	+	+	+	+
9089	AZE	38,9522	48,925	medium	+	+	+	+
9090	AZE	38,9522	48,925	NA	+	+	+	NA
9091	AZE	38,9522	48,925	NA	+	+	+	+
9094	AZE	38,9786	48,5594	NA	+	+	+	NA
9095	AZE	38,9786	48,5594	medium	+	+	+	+
9096	AZE	38,9786	48,5594	NA	+	+	+	NA
9098	AZE	38,9786	48,5594	NA	+	+	+	NA
9099	AZE	38,9786	48,5594	NA	+	+	+	+
9100	GEO	41,8296	46,2831	NA	+	+	+	NA
9102	GEO	41,8296	46,2831	medium	+	+	+	+
9103	GEO	41,8296	46,2831	medium	+	+	+	+
9104	GEO	41,8296	46,2831	medium	+	+	+	+
9105	GEO	41,8296	46,2831	medium	+	+	+	+
9106	GEO	41,8296	46,2831	NA	+	+	+	+
9111	GEO	41,8296	46,2831	medium	+	+	+	+
9113	GEO	41,8296	46,2831	NA	+	+	+	+
9114	GEO	41,8296	46,2831	medium	+	+	+	+
9115	GEO	41,8296	46,2831	medium	+	+	+	+
9119	GEO	41,7942	43,4767	NA	+	+	+	NA
9120	GEO	41,7942	43,4767	NA	+	+	+	NA
9121	GEO	41,7942	43,4767	NA	+	+	+	-
9124	GEO	41,7942	43,4767	NA	+	+	+	NA
9125	ARM	40,1408	44,8203	NA	+	+	+	+
9128	ARM	39,8692	45,3622	medium	+	+	+	+
9130	ARM	39,8692	45,3622	medium	+	+	+	+

Appendix

Accession	Country	Lat	Long	FT promoter	Coding SNP <i>FT</i>	Noncoding SNP <i>FT</i>	TEP <i>FT</i>	<i>FRI</i>
9131	ARM	39,8692	45,3622	medium	+	+	+	+
9133	ARM	39,8692	45,3622	long	+	+	+	+
9134	ARM	39,8692	45,3622	long	+	+	+	+
9298	UK	55,9681	-3,21833	NA	+	+	+	+
9312	UK	57,9	-5,1525	NA	+	-	+	+
9314	UK	57,9672	-3,96722	long	-	+	+	+
9321	SWE	62,8622	18,336	NA	+	+	+	+
9323	SWE	62,8622	18,336	short	+	+	+	+
9330	SWE	56,9591	16,4052	NA	+	+	+	NA
9332	SWE	62,8698	18,381	NA	+	+	+	+
9339	SWE	57,7133	15,0689	medium	+	+	+	+
9343	SWE	57,3089	18,1512	NA	+	+	-	NA
9352	SWE	57,2608	16,3675	long	+	+	-	+
9353	SWE	57,2608	16,3675	long	+	+	-	+
9356	SWE	62,8762	18,1746	NA	+	+	+	NA
9363	SWE	62,9147	18,4045	medium	+	+	+	+
9369	SWE	57,6781	14,9986	NA	+	+	+	+
9371	SWE	63,016	18,3175	medium	-	+	+	-
9380	SWE	55,7488	13,3742	medium	+	+	+	+
9386	SWE	62,806	18,1896	medium	+	+	+	+
9388	SWE	62,806	18,1896	medium	+	+	+	+
9390	SWE	57,3263	15,8979	long	+	+	+	-
9391	SWE	57,3263	15,8979	long	+	+	+	-
9392	SWE	57,3263	15,8979	long	+	+	+	-
9395	SWE	57,5089	15,0105	NA	+	+	+	+
9399	SWE	55,4234	13,9905	medium	+	+	+	+
9402	SWE	57,8765	14,8549	medium	+	+	+	+
9404	SWE	55,7491	13,399	medium	+	+	+	+
9405	SWE	55,7491	13,399	NA	+	+	+	+
9407	SWE	55,7491	13,399	NA	+	+	+	+
9408	SWE	56,047	13,9519	NA	+	+	+	+
9409	SWE	56,0573	14,302	NA	+	+	+	-
9412	SWE	57,2746	16,1494	medium	+	+	+	+
9413	SWE	57,2746	16,1494	medium	+	+	+	+
9416	SWE	57,7215	18,3837	NA	+	+	+	+
9421	SWE	55,9745	14,3997	medium	+	-	+	+
9427	SWE	62,8815	18,4055	medium	+	+	+	+
9433	SWE	62,9513	18,2763	medium	+	+	+	+
9434	SWE	62,8959	18,3659	medium	+	+	+	+
9436	SWE	56,1633	14,6806	short	+	+	+	+
9437	SWE	56,1633	14,6806	medium	+	+	+	+
9442	SWE	55,5678	14,3398	NA	-	+	-	+
9450	SWE	57,2545	18,2109	NA	+	+	+	+
9451	SWE	57,2545	18,2109	short	+	+	+	+
9452	SWE	57,2545	18,2109	medium	+	+	+	+
9453	SWE	57,8009	18,5162	medium	+	+	+	+
9454	SWE	57,8009	18,5162	medium	+	+	+	+

Accession	Country	Lat	Long	FT promoter	Coding SNP FT	Noncoding SNP FT	TEP FT	FRI
9455	SWE	57,8009	18,5162	medium	+	+	+	+
9470	SWE	57,6511	14,8043	medium	+	+	+	+
9476	SWE	55,5796	14,3336	medium	+	+	+	+
9481	SWE	55,4242	13,8484	medium	+	+	+	+
9503	UK	55,8877	-3,21072	medium	+	+	+	-
9507	POR	38,45	-7,5	long	+	-	+	+
9509	POR	39,29	-7,4	NA	+	+	+	+
9510	POR	38,75	-7,59	NA	+	-	+	+
9511	POR	38,53	-8,02	long	+	+	+	+
9512	POR	38,22	-7,84	NA	+	+	+	+
9514	ESP	39,15	-4,54	medium	+	+	+	+
9515	ESP	39,72	-6,89	NA	+	+	+	+
9516	ESP	39,9	-5,09	NA	+	+	+	NA
9517	ESP	42,19	-7,8	NA	+	+	+	+
9518	ESP	39,88	-0,36	medium	+	+	+	+
9519	ESP	41,94	2,64	short	+	+	+	+
9520	ESP	41,7	-3,68	long	+	+	+	+
9521	ESP	41,43	2,13	NA	+	+	+	+
9522	ESP	36,52	-5,27	long	+	+	+	+
9524	ESP	42,52	-0,56	short	+	+	+	+
9525	ESP	42,49	0,54	medium	+	+	+	-
9526	ESP	41,54	2,39	medium	+	+	+	+
9528	ESP	40,94	-1,37	medium	+	+	+	-
9531	ESP	41,21	-4,54	long	+	+	+	+
9532	ESP	42,23	-4,64	NA	+	+	+	+
9533	ESP	41,15	-4,32	long	+	+	+	+
9534	ESP	40,05	-4,65	NA	+	+	+	+
9535	ESP	42,31	3,19	medium	+	+	+	+
9537	ESP	38,07	-6,66	NA	+	+	+	+
9538	ESP	43,12	-8,09	NA	+	+	+	NA
9539	ESP	40,29	-6,67	NA	+	+	+	+
9540	ESP	41,81	2,34	medium	+	+	+	+
9541	ESP	38,26	-5,42	NA	+	+	+	NA
9542	ESP	40,79	-4,05	NA	+	+	+	+
9543	ESP	36,77	-5,39	long	+	+	+	+
9544	ESP	39,4	-5,33	NA	+	+	+	+
9545	ESP	39,4	-5,78	long	+	+	+	+
9546	ESP	40,82	-1,68	NA	+	+	+	+
9547	ESP	41,67	2,62	NA	+	-	+	+
9549	ESP	42,23	-3,69	medium	+	+	+	+
9550	ESP	43,05	-5,37	NA	+	+	+	+
9553	ESP	41,58	-4,71	long	+	+	+	+
9554	ESP	38,86	-3,16	NA	+	+	+	+
9555	ESP	39,58	-3,93	long	+	+	+	+
9556	ESP	39,66	-4,34	medium	+	+	+	+
9557	ESP	42,46	0,7	medium	+	+	+	-
9560	ESP	38,19	-6,24	NA	+	+	+	+

Appendix

Accession	Country	Lat	Long	FT promoter	Coding SNP <i>FT</i>	Noncoding SNP <i>FT</i>	TEP <i>FT</i>	<i>FRI</i>
9562	ESP	41,67	2	medium	+	+	+	+
9563	ESP	40,42	-4,65	NA	+	+	+	NA
9564	ESP	40,45	-1,6	NA	+	+	+	+
9566	ESP	42,44	-4,36	NA	+	+	+	NA
9567	ESP	42,34	1,3	NA	+	+	+	+
9568	ESP	42,76	-0,23	NA	+	+	+	+
9569	ESP	42,87	-6,45	NA	+	+	+	NA
9570	ESP	41,35	1,03	NA	+	+	+	NA
9571	ESP	43,28	-6,01	NA	+	+	+	+
9572	ESP	42,75	-3,05	NA	+	+	+	NA
9573	ESP	41,86	2,99	NA	+	+	+	NA
9574	ESP	38,6	-2,7	long	+	+	+	+
9575	ESP	42,77	-4,21	NA	+	+	+	NA
9577	ESP	42,34	2,17	NA	+	+	+	+
9578	ESP	42,13	-6,7	NA	+	+	+	+
9580	ESP	38,68	-3,57	NA	+	+	+	NA
9582	ESP	41,48	-1,63	long	+	+	+	+
9583	ESP	37,09	-3,38	long	+	+	+	+
9584	ESP	41,19	-3,58	medium	+	+	+	+
9585	ESP	43,4	-7,39	long	+	+	+	+
9586	ESP	41,03	-3,27	medium	+	+	+	+
9587	ESP	41,5	-1,88	NA	+	+	+	+
9588	ESP	42,11	0,6	NA	+	+	+	+
9589	ESP	41,6	-2,83	medium	+	+	+	+
9590	ESP	43,37	-5,49	NA	+	+	+	+
9593	ESP	42,26	-2,99	NA	+	+	+	+
9594	ESP	42,04	1,01	medium	+	+	+	+
9597	ESP	42,31	-2,53	NA	+	+	+	+
9598	ESP	41,88	-6,51	NA	+	+	+	+
9599	ESP	42,8	-5,77	medium	+	+	+	+
9600	ESP	39,85	-6,04	long	+	+	+	+
9601	ESP	41,85	-1,88	medium	+	+	+	+
9602	ESP	40,5	-3,96	medium	+	+	+	+
9603	ESP	42,83	-4,72	NA	+	+	+	NA
9604	ESP	42,1	-2,35	NA	+	+	+	NA
9605	ESP	40,55	-4,19	NA	+	+	+	NA
9606	MAR	31,48	-7,45	NA	+	+	+	+
9607	RUS	53,05	52,15	NA	+	+	+	+
9608	RUS	51,37	59,44	medium	+	+	+	+
9609	RUS	51,41	59,98	NA	+	+	+	+
9610	RUS	53,04	51,96	NA	+	+	+	-
9611	RUS	53,04	51,9	medium	+	+	+	+
9612	RUS	53,04	51,94	NA	+	+	+	+
9613	RUS	55,36	61,41	medium	+	+	+	+
9614	RUS	55,53	65,33	NA	+	+	+	NA
9615	RUS	52,99	52,16	medium	+	+	+	+
9616	RUS	53,06	51,96	NA	+	+	+	-

Accession	Country	Lat	Long	FT promoter	Coding SNP FT	Noncoding SNP FT	TEP FT	FRI
9617	RUS	51,37	59,44	NA	+	+	+	-
9618	RUS	55,61	65,08	NA	+	+	+	NA
9619	RUS	51,84	79,48	medium	+	+	+	+
9620	RUS	51,82	79,48	NA	+	+	+	+
9621	RUS	51,84	79,46	medium	+	+	+	+
9623	RUS	53,6	79,39	NA	+	+	+	NA
9624	RUS	53,6	79,37	NA	+	+	+	+
9625	RUS	51,31	82,59	medium	+	+	+	+
9626	RUS	51,36	82,59	medium	+	+	+	+
9627	RUS	51,32	82,55	NA	+	+	+	-
9628	RUS	51,33	82,54	NA	+	+	+	+
9629	RUS	51,35	82,18	medium	+	+	+	+
9630	RUS	51,34	82,16	medium	+	+	+	+
9631	RUS	51,65	80,79	NA	+	+	+	+
9632	RUS	51,67	80,82	NA	+	+	+	+
9633	RUS	51,63	80,83	NA	+	+	+	-
9634	RUS	54,13	81,31	NA	+	+	+	+
9635	RUS	51,87	80,6	NA	+	+	+	-
9636	RUS	51,75	80,82	NA	+	+	+	-
9637	RUS	51,77	80,85	NA	+	+	+	+
9638	RUS	51,73	80,86	NA	+	+	+	+
9639	RUS	53,82	80,31	medium	+	+	+	-
9640	RUS	51,87	80,06	NA	+	+	+	+
9641	RUS	51,9	80,06	NA	+	+	+	+
9642	RUS	51,84	80,06	NA	+	+	+	-
9643	RUS	52,1	79,31	medium	+	+	+	+
9644	CRO	45,07	18,72	long	+	+	+	-
9645	CRO	45,17	18,7	NA	+	+	+	+
9646	ITA	39,126899	16,170188	medium	+	+	+	-
9647	ITA	40,37	16,77	medium	+	+	+	+
9648	ITA	39,48	16,28	NA	+	+	+	NA
9649	ITA	39,13	16,17	NA	+	+	+	+
9650	ITA	39,6	16,51	NA	+	+	+	NA
9651	ITA	40,68	14,87	medium	+	+	+	-
9652	ITA	41,36	13,4	NA	+	+	+	NA
9653	ITA	38,44	16,13	NA	+	+	+	+
9654	ITA	41,41	13,77	NA	+	+	+	NA
9655	ITA	38,92	16,47	NA	+	+	+	+
9656	ITA	40,64	17,31	NA	+	+	+	+
9657	ITA	38,45	16,04	NA	+	+	+	+
9658	ITA	38,97	16,34	NA	+	+	+	+
9659	ITA	41,18	14,18	NA	+	+	+	+
9660	ITA	40,84	14,57	short	+	+	+	+
9661	ITA	39,58	16,21	NA	+	+	+	+
9662	ITA	38,47	16,47	NA	+	+	+	NA
9664	ITA	46,36	11,28	long	+	+	+	+
9665	ITA	46,36	11,28	NA	+	+	+	+

Appendix

Accession	Country	Lat	Long	FT promoter	Coding SNP <i>FT</i>	Noncoding SNP <i>FT</i>	TEP <i>FT</i>	<i>FRI</i>
9666	ITA	46,36	11,28	NA	+	+	+	+
9667	ITA	46,36	11,28	medium	+	+	+	+
9668	ITA	46,37	11,28	medium	+	+	+	-
9669	ITA	46,37	11,28	NA	+	+	+	+
9670	ITA	46,37	11,28	medium	+	+	+	-
9671	ITA	46,37	11,28	NA	+	+	+	+
9672	ITA	46,37	11,28	medium	+	+	+	+
9673	ITA	46,37	11,28	medium	+	+	+	+
9674	ITA	46,37	11,29	NA	+	+	+	NA
9675	ITA	46,37	11,29	NA	+	+	+	NA
9676	ITA	46,37	11,29	NA	+	+	+	-
9677	ITA	46,37	11,29	NA	+	+	+	+
9678	ITA	46,37	11,29	NA	+	+	+	+
9679	ITA	46,34	11,29	NA	+	+	+	+
9680	ITA	46,34	11,29	medium	+	+	+	-
9681	ITA	46,34	11,29	NA	+	+	+	+
9682	ITA	46,34	11,29	NA	+	+	+	+
9683	ITA	46,34	11,29	NA	+	+	+	+
9684	ITA	46,34	11,29	NA	+	+	+	+
9685	ITA	46,34	11,29	NA	+	+	+	+
9686	ITA	46,34	11,29	medium	+	+	+	+
9687	ITA	46,34	11,29	NA	+	+	+	+
9688	ITA	46,34	11,29	NA	+	+	+	NA
9689	ITA	46,34	11,29	medium	+	+	+	+
9690	ITA	46,34	11,29	medium	+	+	+	+
9692	ITA	46,34	11,29	medium	+	+	+	+
9693	ITA	46,34	11,29	medium	+	+	+	+
9694	ITA	46,34	11,29	medium	+	+	+	+
9695	ITA	46,34	11,29	NA	+	+	+	-
9696	ITA	46,34	11,29	NA	+	+	+	+
9697	BUL	41,62	23,94	medium	+	+	+	+
9698	BUL	41,57	23,85	long	+	+	+	+
9699	BUL	41,37	23,14	medium	+	+	+	+
9700	BUL	42,32	23,1	medium	+	+	+	+
9701	BUL	43,7	25,91	medium	+	+	+	-
9702	BUL	41,38	23,14	NA	+	+	+	NA
9703	BUL	41,53	23,39	medium	+	+	+	+
9704	BUL	41,53	23,39	medium	+	+	+	-
9705	BUL	41,5	23,33	NA	+	+	+	+
9706	BUL	41,64	24,18	medium	+	+	+	-
9707	BUL	41,57	24,84	NA	+	+	+	+
9708	BUL	41,62	25,35	short	+	+	+	+
9709	BUL	41,85	23,13	medium	+	+	+	+
9710	BUL	41,85	23,13	long	+	+	+	+
9711	BUL	42,32	23,1	medium	+	+	+	+
9712	BUL	42,32	23,1	NA	+	+	+	NA
9713	BUL	42,49	25,61	NA	+	+	+	-

Accession	Country	Lat	Long	FT promoter	Coding SNP FT	Noncoding SNP FT	TEP FT	FRI
9714	BUL	41,84	25,75	NA	+	+	+	+
9715	BUL	41,99	25,57	NA	+	+	+	NA
9716	BUL	41,54	24,98	NA	+	+	+	+
9717	BUL	41,66	25,47	NA	+	+	+	+
9718	BUL	41,55	24,75	NA	+	+	+	+
9719	BUL	41,83	25,69	long	+	+	+	-
9720	BUL	41,77	25,68	short	+	+	+	+
9721	BUL	42,72	25,33	medium	+	+	+	-
9722	BUL	41,71	24,41	NA	+	+	+	+
9723	BUL	41,42	23,67	NA	+	+	+	-
9724	BUL	41,54	24,98	NA	+	+	+	NA
9725	GRC	37,6	23,08	NA	+	+	+	+
9726	GRC	37,07	22,04	NA	+	+	+	-
9727	GRC	37,63	21,62	NA	+	+	+	+
9728	SVK	48,46	18,9	long	+	+	+	+
9729	SVK	48,46	18,9	medium	+	+	+	+
9730	SVK	48,47	18,94	NA	+	+	+	+
9731	SVK	48,46	18,9	medium	+	+	+	+
9732	SVK	48,47	18,96	medium	+	+	+	+
9733	SVK	48,47	18,94	NA	+	+	+	+
9734	SVK	48,47	18,94	NA	+	+	+	NA
9735	SVK	48,47	18,94	NA	+	+	+	+
9737	ROU	45,95	22,62	NA	+	+	+	-
9739	ROU	46,01	22,33	long	-	+	+	+
9740	ROU	46,16	21,43	NA	+	+	+	NA
9741	ROU	45,84	23,16	long	+	+	+	+
9742	ROU	44,68	25,17	NA	+	+	+	NA
9745	UZB	41,45	70,05	NA	+	+	+	-
9746	SRB	43,71	22,3	NA	+	+	+	NA
9747	SRB	44,38	21,22	NA	+	+	+	+
9749	SRB	43,54	22,29	NA	+	+	+	-
9750	SRB	43	22,65	NA	+	+	+	NA
9751	SRB	44,91	19,99	NA	+	+	+	NA
9752	SRB	44	22,07	NA	+	+	+	NA
9753	SRB	44,91	19,99	NA	+	+	+	NA
9755	SRB	44,56	22,56	medium	+	+	+	+
9756	SRB	44,3	21,08	medium	+	+	+	+
9757	SRB	44,3	21,08	medium	+	+	+	+
9758	CHN	47,75	88,4	medium	+	+	+	+
9759	IRN	37,47	49,47	medium	+	+	+	+
9760	FRA	48,81	1,66	NA	+	+	+	NA
9761	LBN	33,92	35,7	long	+	+	+	-
9766	KGZ	42,26	74,16	medium	+	+	+	-
9767	LBN	34,25	35,92	NA	+	+	+	NA
9768	GER	48,57	9,16	NA	+	+	+	-
9769	GER	48,55	8,99	short	+	+	+	-
9770	GER	48,53	9,01	long	+	+	+	-

Appendix

Accession	Country	Lat	Long	FT promoter	Coding SNP <i>FT</i>	Noncoding SNP <i>FT</i>	TEP <i>FT</i>	<i>FRI</i>
9771	GER	48,56	9,11	long	+	+	+	-
9772	GER	48,41	8,85	medium	+	+	+	+
9773	GER	48,52	8,92	NA	+	+	+	NA
9774	GER	48,59	9,22	medium	+	+	+	-
9775	GER	48,41	8,79	NA	+	-	+	+
9776	GER	48,43	8,79	medium	+	+	+	-
9777	GER	48,57	9,17	long	+	+	+	-
9778	GER	48,41	8,84	long	+	+	+	NA
9779	GER	48,5	8,78	NA	+	-	+	+
9780	GER	48,43	8,79	long	+	+	+	NA
9781	GER	48,52	9,11	medium	+	+	+	-
9782	GER	48,53	9,09	long	+	+	+	-
9784	GER	48,5	8,8	long	+	+	+	-
9785	GER	48,54	9,02	medium	+	+	+	-
9786	GER	48,54	9,01	long	+	+	+	-
9787	GER	48,5	9	NA	+	+	+	NA
9788	GER	48,53	9,01	long	+	+	+	-
9789	GER	48,39	8,96	long	+	+	+	-
9790	GER	48,58	9,18	long	+	+	+	-
9791	GER	48,6	9,2	long	-	+	+	-
9792	GER	48,54	9,09	long	+	+	+	-
9793	GER	48,57	9,16	long	+	+	+	-
9794	GER	48,52	9,08	long	+	+	+	+
9795	GER	48,5	9,11	medium	+	+	+	+
9796	GER	48,41	8,84	medium	+	+	+	-
9797	GER	48,54	9,02	long	+	+	+	-
9798	GER	48,54	9,01	medium	+	+	+	-
9799	GER	48,39	8,85	medium	+	+	+	-
9800	GER	48,54	9,01	long	+	+	+	-
9801	GER	48,54	9,01	long	+	+	+	-
9802	GER	48,51	9,11	NA	+	+	+	-
9803	GER	48,42	8,76	NA	+	+	+	-
9804	GER	48,45	8,87	long	+	+	+	-
9805	GER	48,54	9,09	long	+	+	+	-
9806	GER	48,56	9,16	long	-	+	+	-
9807	GER	48,6	9,22	long	+	+	+	-
9808	GER	48,52	9,08	NA	+	+	+	-
9809	GER	48,52	9,05	NA	+	+	+	+
9810	GER	48,53	9,07	long	+	+	+	-
9811	GER	48,52	9,05	long	+	+	+	-
9812	GER	48,52	9,03	NA	+	+	+	-
9813	GER	48,4	8,77	NA	+	+	+	+
9814	GER	48,42	8,79	long	+	+	+	-
9815	GER	48,54	9,02	NA	+	+	+	+
9816	GER	48,55	9,06	long	+	+	+	+
9817	ESP	39,84	-6,6	NA	+	+	+	+
9818	ESP	41,14	-0,25	NA	+	+	+	NA

Accession	Country	Lat	Long	FT promoter	Coding SNP FT	Noncoding SNP FT	TEP FT	FRI
9819	ESP	42,35	-3,03	NA	+	+	+	+
9820	ESP	41	-4,71	medium	+	+	+	+
9821	ESP	41,81	2,49	medium	+	+	+	+
9822	ESP	40,52	-4,02	NA	+	+	+	+
9823	ESP	43,34	-5,84	NA	+	+	+	+
9825	ESP	40,4	-3,88	NA	+	+	+	+
9826	ESP	42,49	-6,71	NA	+	+	+	+
9827	ESP	42,78	0,69	medium	+	-	+	+
9828	ESP	42,5	-6,15	NA	+	+	+	+
9829	ESP	40,43	-4,75	NA	+	+	+	NA
9832	ESP	40,54	-3,69	long	+	+	+	+
9833	ESP	40,38	-4,21	NA	+	+	+	+
9834	ESP	40,51	-3,9	medium	+	+	+	+
9836	ESP	41,25	-1,32	NA	+	+	+	+
9837	ESP	37,94	-5,6	NA	+	+	+	+
9840	ESP	41,13	-1,43	NA	+	+	+	+
9841	ESP	40,59	-4,15	NA	+	+	+	+
9842	ESP	39,31	-3,89	NA	+	+	+	NA
9843	ESP	40,53	-3,92	long	+	+	+	+
9844	ESP	42,27	0,19	NA	+	+	+	+
9845	ESP	40,48	-3,96	NA	+	+	+	+
9846	ESP	42,31	-3,02	medium	+	+	+	+
9847	ESP	43,31	-5,7	NA	+	+	+	+
9848	ESP	40,11	-5,77	NA	+	+	+	+
9850	ESP	42,86	-0,7	NA	+	+	+	+
9851	ESP	42,96	-6,1	medium	+	+	+	+
9852	ESP	40,46	-3,75	NA	+	+	+	+
9853	ESP	43,33	-5,91	NA	+	+	+	+
9854	ESP	43,36	-5,88	medium	+	+	+	-
9855	ESP	40,57	-3,89	long	+	+	+	+
9856	ESP	40,51	-4	NA	+	+	+	NA
9857	ESP	40,33	-3,8	NA	+	+	+	NA
9859	ESP	40,5	-3,88	NA	+	+	+	NA
9860	ESP	42,24	-2,62	NA	+	+	+	-
9861	ESP	40,72	-3,21	NA	+	+	+	NA
9862	ESP	40,45	-3,67	NA	+	+	+	NA
9863	ESP	43,33	-5,87	NA	+	+	+	NA
9864	ESP	41,76	2,69	NA	+	+	+	+
9865	ESP	38,88	-3,53	NA	+	+	+	NA
9866	ESP	41,89	-2,79	medium	+	+	+	+
9867	ESP	40,94	-3,22	NA	+	+	+	+
9868	ESP	41,78	2,37	medium	+	+	+	+
9869	ESP	36,76	-5,28	long	+	+	+	+
9870	ESP	41,91	0,17	long	+	+	+	+
9871	ESP	40,75	-3,99	NA	+	+	+	+
9872	ESP	40,49	-4,11	NA	+	+	+	NA
9873	ESP	37,94	-5,45	NA	+	+	+	+

Appendix

Accession	Country	Lat	Long	FT promoter	Coding SNP <i>FT</i>	Noncoding SNP <i>FT</i>	TEP <i>FT</i>	<i>FRI</i>
9874	ESP	42,34	-3	NA	+	+	+	+
9875	ESP	43,38	-5,87	NA	+	+	+	+
9876	ESP	41,34	0,99	NA	+	+	+	+
9878	ESP	40,78	-3,62	long	+	-	+	+
9879	ESP	37,6	-1,12	long	+	+	+	+
9880	ESP	42,72	-3,44	NA	+	+	+	+
9882	ESP	40,46	-4,26	NA	+	+	+	+
9883	ESP	42,1	-2,56	short	+	+	+	-
9884	ESP	42,25	-3,04	NA	+	+	+	NA
9885	ESP	41,14	-3,68	medium	+	+	+	+
9886	ESP	42,38	1,73	NA	+	+	+	-
9887	ESP	40,4	-4,77	long	+	+	+	+
9888	ESP	40,93	-3,31	long	+	+	+	+
9889	ESP	40,86	-3,87	NA	+	+	+	NA
9891	ESP	41,93	2,92	medium	+	+	+	+
9893	ESP	37,96	-5,37	NA	+	+	+	NA
9895	ESP	41,78	2,57	medium	+	+	+	+
9896	ESP	38,47	-3,74	NA	+	+	+	NA
9898	ESP	41,14	-3,58	NA	+	+	+	NA
9899	ESP	42,54	0,84	medium	+	+	+	-
9900	ESP	37,38	-6,01	NA	+	+	+	+
9901	ESP	42,27	-2,98	NA	+	+	+	+
9902	ESP	40,71	-3,24	medium	+	+	+	+
9903	ESP	42,31	-3,1	medium	+	+	+	+
9904	ESP	40,95	-3,31	NA	+	+	+	+
9905	ESP	40,76	-4,01	long	+	+	+	+
9907	FRA	50,86	3,6	NA	+	+	+	NA
9908	FRA	50,72	3,47	long	+	+	+	+
9909	FRA	50,59	3,3	NA	+	-	+	+
9910	FRA	50,68	3,52	medium	+	+	+	-
9911	FRA	47,16	4,28	long	-	+	+	+
9913	FRA	47,24	4,43	NA	+	+	+	NA
9914	FRA	47,58	5,33	medium	+	+	+	+
9915	FRA	47,45	3,94	medium	+	+	+	+
9916	FRA	47,1	4,22	NA	+	+	+	NA
9917	FRA	46,69	4,34	medium	+	-	+	+
9918	FRA	47,43	5,21	NA	+	+	+	+
9919	FRA	48,85	4,45	NA	+	+	+	NA
9920	FRA	48,54	4,32	NA	+	+	+	+
9921	FRA	48,57	4,41	NA	+	+	+	-
9922	FRA	48,52	4,7	NA	+	+	+	NA
9923	FRA	48,58	4,46	NA	+	+	+	NA
9925	FRA	48,91	4,52	medium	+	+	+	-
9926	FRA	48,86	4,1	NA	+	+	+	-
9927	FRA	44,05	3,69	medium	+	+	+	+
9928	FRA	44,12	3,77	NA	+	-	+	+
9929	FRA	43,92	3,71	long	+	+	+	+

Accession	Country	Lat	Long	FT promoter	Coding SNP FT	Noncoding SNP FT	TEP FT	FRI
9930	FRA	43,91	4,14	NA	+	+	+	-
9931	FRA	43,98	4,31	NA	+	+	+	NA
9934	FRA	44,07	4,08	NA	+	+	+	NA
9935	FRA	50,6	2,93	NA	+	+	+	-
9936	FRA	50,47	3,46	NA	+	+	+	NA
9937	FRA	50,79	2,69	medium	+	+	+	+
9938	FRA	50,65	2,99	NA	+	+	+	+
9939	MAR	31,48	-7,45	NA	+	+	+	-
9940	MAR	31,47	-7,42	NA	+	+	+	NA
9941	POR	40,92	-8,54	NA	+	+	+	+
9942	ESP	41,32	-1,34	short	+	+	+	+
9943	ESP	39,73	-5,74	short	+	+	+	+
9944	ESP	36,83	-6,36	long	+	+	+	+
9946	ESP	38,92	-6,34	NA	+	+	+	+
9947	ESP	40,74	-3,9	long	+	+	+	+
9948	ESP	41,05	-3,54	NA	+	+	+	+
9949	ESP	42,69	-6,93	NA	+	+	+	+
9950	ESP	42,63	0,76	NA	+	+	+	+
9951	RUS	51,3333	82,5667	NA	+	+	+	-
9952	RUS	51,32	82,55	NA	+	+	+	+
9953	RUS	51,33	82,19	NA	+	+	+	+
9954	RUS	51,65	80,82	NA	+	+	+	NA
9955	RUS	54,09	60,46	NA	+	+	+	-
9956	RUS	54,06	60,48	NA	+	+	+	+
9957	RUS	53,04	51,75	NA	+	+	+	-
9958	RUS	53,33	49,48	NA	+	+	+	+
9959	RUS	53,33	49,48	NA	+	+	+	+
9960	RUS	51,31	57,56	NA	+	+	+	-
9961	RUS	53,09	52	NA	+	+	+	NA
9962	ITA	40,57	15,32	NA	+	+	+	+
9963	ITA	39,18	16,26	short	+	+	+	+
9964	ITA	38,36	16,23	short	+	+	-	-
9965	ITA	38,38	16,22	NA	+	+	+	+
9966	ITA	40,28	15,65	NA	+	+	+	+
9967	ITA	39,83	16,17	short	+	+	-	NA
9968	ITA	39,27	16,27	short	+	+	+	+
9969	ITA	40,18	16,45	NA	+	+	+	+
9970	ITA	46,37	11,24	NA	+	+	+	+
9971	ITA	46,51	11,33	long	+	+	+	+
9972	ITA	46,51	11,33	long	+	+	+	+
9973	ITA	46,36	11,28	NA	+	+	+	-
9974	ITA	46,34	11,29	NA	+	+	+	NA
9975	ITA	46,34	11,29	NA	+	+	+	+
9976	ITA	46,25	11,17	NA	-	+	+	+
9977	ITA	46,63	10,82	short	+	+	+	NA
9978	ITA	46,63	10,82	short	+	+	+	-
9979	ITA	46,36	11,23	long	+	+	+	+

Appendix

Accession	Country	Lat	Long	FT promoter	Coding SNP FT	Noncoding SNP FT	TEP FT	FRI
9980	ITA	38,62	16,17	short	+	+	-	-
9981	ITA	38,76	16,24	short	+	+	+	-
9982	ITA	39,01	16,47	short	+	+	+	+
9985	BUL	41,43	23,65	NA	+	+	+	+
9986	BUL	41,59	25,2	NA	+	+	+	+
9987	BUL	41,43	23,5	NA	+	+	+	+
9988	GEO	41,7942	43,4767	NA	+	+	+	+
9989	GEO	41,7942	43,4767	NA	+	+	+	NA
9990	GEO	41,8296	46,2831	NA	+	+	+	NA
9991	GEO	41,2381	46,3728	NA	+	+	+	-
9992	TUR	38,3011	42,2239	NA	+	+	+	NA
9994	GER	48,43	8,77	NA	+	+	+	NA
9995	GER	48,14	9,4	NA	+	+	+	+
9996	GER	48,52	8,8	long	+	+	+	-
9997	GER	48,56	9,16	long	+	+	+	NA
9999	GER	48,53	9,06	long	+	+	+	-
10000	GER	48,53	9,05	NA	+	+	+	NA
10001	GER	48,52	9,05	NA	+	+	+	NA
10002	GER	48,53	9,04	NA	+	+	+	-
10003	GER	48,6	9,19	NA	+	+	+	NA
10004	ROU	44,46	25,74	NA	+	+	+	+
10005	ROU	46,11	21,95	NA	+	+	+	+
10006	UKR	44,6419	34,3814	NA	+	+	+	-
10007	UKR	50,3553	29,3244	NA	+	+	+	NA
10008	UZB	41,45	70,05	NA	+	+	+	+
10009	UZB	41,45	70,05	NA	+	+	+	-
10010	UZB	41,45	70,05	NA	+	+	+	+
10011	ARM	39,8692	45,3622	NA	+	+	+	+
10012	AZE	38,9786	48,5594	NA	+	+	+	-
10013	AZE	38,7406	48,6131	NA	+	+	+	-
10014	AZE	38,6536	48,7992	NA	+	+	+	-
10015	AFG	37,29	71,3	NA	+	+	+	NA
10016	SRV	44,9444	21,1828	NA	+	+	+	NA
10017	SRB	44,34	21,46	NA	+	+	+	+
10018	SRB	44,84	20,16	NA	+	+	-	
10018	SRB	44,84	20,16	NA	+	+	+	+
10019	KAZ	48,67	54,93	NA	+	+	+	NA
10020	CZE	49,17	16,5	medium	+	+	+	+
10021	GER	52,1444	9,37827	NA	+	+	+	NA
10024	TZA	NA	NA	NA	+	+	+	NA
10025	GER	50,01	8,67	NA	+	+	+	NA
10026	USA	47,6479	-122,305	NA	+	+	+	NA
15591	AUT	48,3315333	14,72665	NA	+	+	-	NA
15592	AUT	48,3314667	14,7158667	NA	+	+	-	NA
100000	LTU	54,6833	25,3167	NA	+	+	+	NA

ID	URL
88	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945435
108	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945436
139	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945437
159	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945438
265	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945439
350	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945440
403	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945442
410	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945443
424	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945444
428	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945445
430	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR492/SRR492273/SRR492273.sra
470	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945447
476	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945448
484	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945449
504	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945450
506	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945451
531	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945452
544	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945453
546	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945454
628	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945455
680	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945457
681	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945458
685	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945459
687	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945460
728	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945461
742	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945462
763	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR492/SRR492291/SRR492291.sra
765	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR492/SRR492381/SRR492381.sra
766	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR492/SRR492244/SRR492244.sra
768	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR492/SRR492410/SRR492410.sra
772	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR492/SRR492330/SRR492330.sra
801	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945468
853	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945469
854	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945470
867	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945471
868	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945472
870	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945473
915	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945474
932	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945475
991	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR519/SRR519714/SRR519714.sra
992	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR519/SRR519715/SRR519715.sra
997	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR519/SRR519716/SRR519716.sra
1006	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945480
1070	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945485
1166	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945487
1313	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945490
1612	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945493
1622	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945494
1651	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945495
1652	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945496
1676	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945497
1684	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945498
1739	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945499
1741	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945500
1756	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945501
1757	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945502
1793	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945503
1797	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945504
1819	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945505
1820	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945506
1829	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945507
1834	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945508
1835	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945509
1851	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945510
1852	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945511

Supplementary Table 6: see page 134

Appendix

ID	URL
1853	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945512
1872	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945513
1890	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945514
1925	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945515
1942	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945516
1943	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945517
1954	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945518
2016	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945519
2017	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945520
2031	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945521
2053	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945522
2057	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945523
2081	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945524
2091	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945525
2106	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945526
2108	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945527
2141	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945528
2159	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945529
2166	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945530
2171	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945531
2191	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945532
2202	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945533
2212	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945534
2239	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945535
2240	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945536
2276	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945537
2278	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945538
2285	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945539
2286	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945540
2317	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945541
2370	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945542
2412	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945543
4779	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945544
4807	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945545
4826	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945546
4840	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945547
4857	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945548
4884	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945549
4900	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945550
4939	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945551
4958	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945552
5023	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945553
5104	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945554
5151	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945555
5165	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945556
5210	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945557
5236	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945558
5249	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945559
5253	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945560
5276	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945561
5279	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945562
5349	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945563
5353	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945564
5395	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945565
5486	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945566
5577	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945567
5644	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945568
5651	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945569
5717	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945570
5718	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945571
5720	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945572
5726	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945573
5741	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945574
5748	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945575
5757	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945576

Supplementary Table 6: see page 134

ID	URL
5768	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945577
5772	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945578
5776	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945579
5779	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945580
5784	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945581
5798	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945582
5800	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945583
5811	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945584
5822	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945585
5836	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945589
5837	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR492/SRR492221/SRR492221.sra
5860	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945592
5874	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945595
5890	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945596
5893	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945597
5907	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945598
5921	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945599
5950	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945600
5984	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945601
5993	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945602
6008	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR492/SRR492250/SRR492250.sra
6010	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945605
6022	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945614
6077	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945636
6087	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945639
6091	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945642
6095	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945645
6096	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945646
6101	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945651
6109	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR519/SRR519547/SRR519547.sra
6111	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR519/SRR519548/SRR519548.sra
6112	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR519/SRR519549/SRR519549.sra
6113	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR519/SRR519550/SRR519550.sra
6114	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR519/SRR519551/SRR519551.sra
6115	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR519/SRR519552/SRR519552.sra
6118	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR519/SRR519553/SRR519553.sra
6119	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945665
6122	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR519/SRR519554/SRR519554.sra
6123	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945667
6124	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR519/SRR519555/SRR519555.sra
6125	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR519/SRR519556/SRR519556.sra
6126	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR519/SRR519557/SRR519557.sra
6127	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR519/SRR519558/SRR519558.sra
6128	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR519/SRR519559/SRR519559.sra
6131	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR519/SRR519560/SRR519560.sra
6132	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR519/SRR519561/SRR519561.sra
6133	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR519/SRR519562/SRR519562.sra
6134	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945675
6136	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR519/SRR519563/SRR519563.sra
6137	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR519/SRR519564/SRR519564.sra
6138	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR519/SRR519565/SRR519565.sra
6140	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR519/SRR519566/SRR519566.sra
6141	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945680
6141	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR519/SRR519567/SRR519567.sra
6142	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR519/SRR519568/SRR519568.sra
6145	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR519/SRR519569/SRR519569.sra
6147	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR519/SRR519570/SRR519570.sra
6148	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR519/SRR519571/SRR519571.sra
6149	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR519/SRR519572/SRR519572.sra
6150	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR519/SRR519573/SRR519573.sra
6151	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR519/SRR519574/SRR519574.sra
6153	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR519/SRR519575/SRR519575.sra
6154	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR519/SRR519576/SRR519576.sra
6163	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR519/SRR519577/SRR519577.sra
6166	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR519/SRR519578/SRR519578.sra

Supplementary Table 6: see page 134

Appendix

ID	URL
6169	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR519/SRR519579/SRR519579.sra
6171	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR519/SRR519580/SRR519580.sra
6172	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR519/SRR519581/SRR519581.sra
6173	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR519/SRR519582/SRR519582.sra
6174	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR519/SRR519583/SRR519583.sra
6177	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR519/SRR519584/SRR519584.sra
6180	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR519/SRR519585/SRR519585.sra
6184	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR519/SRR519586/SRR519586.sra
6188	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR519/SRR519587/SRR519587.sra
6189	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR519/SRR519588/SRR519588.sra
6191	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR519/SRR519589/SRR519589.sra
6192	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR519/SRR519590/SRR519590.sra
6193	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR519/SRR519591/SRR519591.sra
6194	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR519/SRR519592/SRR519592.sra
6195	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR519/SRR519593/SRR519593.sra
6198	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR519/SRR519594/SRR519594.sra
6201	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR519/SRR519595/SRR519595.sra
6202	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR519/SRR519596/SRR519596.sra
6203	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR519/SRR519597/SRR519597.sra
6209	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR519/SRR519598/SRR519598.sra
6210	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR519/SRR519599/SRR519599.sra
6214	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945711
6216	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR519/SRR519600/SRR519600.sra
6217	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR519/SRR519601/SRR519601.sra
6218	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR519/SRR519602/SRR519602.sra
6220	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR519/SRR519603/SRR519603.sra
6221	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR519/SRR519604/SRR519604.sra
6231	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR519/SRR519605/SRR519605.sra
6235	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR519/SRR519606/SRR519606.sra
6236	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR519/SRR519607/SRR519607.sra
6237	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR519/SRR519608/SRR519608.sra
6238	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR519/SRR519609/SRR519609.sra
6240	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR519/SRR519610/SRR519610.sra
6241	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR519/SRR519611/SRR519611.sra
6242	ftp://ftp-trace

Supplementary Table 6: see page 134

ID	URL
6904	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR492/SRR492223/SRR492223.sra
6907	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR492/SRR492233/SRR492233.sra
6908	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR492/SRR492234/SRR492234.sra
6909	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR519/SRR519624/SRR519624.sra
6911	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR492/SRR492239/SRR492239.sra
6913	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR519/SRR519625/SRR519625.sra
6915	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR492/SRR492252/SRR492252.sra
6917	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR519/SRR519626/SRR519626.sra
6918	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR519/SRR519627/SRR519627.sra
6919	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR492/SRR492266/SRR492266.sra
6920	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR492/SRR492271/SRR492271.sra
6922	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR492/SRR492274/SRR492274.sra
6923	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR492/SRR492282/SRR492282.sra
6924	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR492/SRR492283/SRR492283.sra
6926	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR492/SRR492298/SRR492298.sra
6927	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945769
6929	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR492/SRR492304/SRR492304.sra
6931	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR492/SRR492308/SRR492308.sra
6933	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR492/SRR492316/SRR492316.sra
6938	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR492/SRR492324/SRR492324.sra
6940	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR492/SRR492326/SRR492326.sra
6943	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR492/SRR492331/SRR492331.sra
6944	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR492/SRR492332/SRR492332.sra
6945	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR492/SRR492333/SRR492333.sra
6951	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR492/SRR492349/SRR492349.sra
6956	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR492/SRR492350/SRR492350.sra
6957	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR492/SRR492351/SRR492351.sra
6958	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR492/SRR492353/SRR492353.sra
6959	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR492/SRR492359/SRR492359.sra
6960	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945784
6961	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR492/SRR492369/SRR492369.sra
6963	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR492/SRR492375/SRR492375.sra
6964	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR519/SRR519628/SRR519628.sra
6965	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR519/SRR519629/SRR519629.sra
6966	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR4

Supplementary Table 6: see page 134

Appendix

ID	URL
7033	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945817
7036	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR492/SRR492226/SRR492226.sra
7061	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR492/SRR492229/SRR492229.sra
7062	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR492/SRR492228/SRR492228.sra
7064	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR492/SRR492235/SRR492235.sra
7068	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR492/SRR492230/SRR492230.sra
7071	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR492/SRR492231/SRR492231.sra
7072	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR492/SRR492232/SRR492232.sra
7077	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR492/SRR492237/SRR492237.sra
7081	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR492/SRR492236/SRR492236.sra
7092	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR492/SRR492238/SRR492238.sra
7096	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR492/SRR492243/SRR492243.sra
7102	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR492/SRR492245/SRR492245.sra
7103	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR492/SRR492247/SRR492247.sra
7106	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR492/SRR492246/SRR492246.sra
7107	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR492/SRR492251/SRR492251.sra
7109	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR492/SRR492255/SRR492255.sra
7117	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR492/SRR492254/SRR492254.sra
7119	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR492/SRR492257/SRR492257.sra
7120	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR492/SRR492258/SRR492258.sra
7125	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR492/SRR492259/SRR492259.sra
7126	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR492/SRR492260/SRR492260.sra
7127	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR492/SRR492261/SRR492261.sra
7130	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR492/SRR492262/SRR492262.sra
7133	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR492/SRR492265/SRR492265.sra
7138	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR492/SRR492264/SRR492264.sra
7143	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR492/SRR492268/SRR492268.sra
7147	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR492/SRR492269/SRR492269.sra
7158	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945850
7160	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR492/SRR492253/SRR492253.sra
7162	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR492/SRR492284/SRR492284.sra
7163	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR492/SRR492276/SRR492276.sra
7164	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR492/SRR492277/SRR492277.sra
7165	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR492/SRR492280/SRR

Supplementary Table 6: see page 134

ID	URL
7316	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR492/SRR492360/SRR492360.sra
7319	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR492/SRR492364/SRR492364.sra
7320	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR492/SRR492365/SRR492365.sra
7323	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR492/SRR492848/SRR492848.sra
7327	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR492/SRR492372/SRR492372.sra
7332	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR492/SRR492370/SRR492370.sra
7333	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR492/SRR492371/SRR492371.sra
7337	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR492/SRR492374/SRR492374.sra
7342	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR492/SRR492380/SRR492380.sra
7343	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR492/SRR492376/SRR492376.sra
7344	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR492/SRR492373/SRR492373.sra
7346	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945906
7347	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR492/SRR492379/SRR492379.sra
7349	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR492/SRR492382/SRR492382.sra
7350	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945909
7353	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR492/SRR492385/SRR492385.sra
7354	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR492/SRR492386/SRR492386.sra
7356	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR492/SRR492387/SRR492387.sra
7358	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945913
7359	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945914
7372	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR492/SRR492390/SRR492390.sra
7377	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR492/SRR492392/SRR492392.sra
7378	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR492/SRR492394/SRR492394.sra
7382	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR492/SRR492397/SRR492397.sra
7383	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR492/SRR492398/SRR492398.sra
7384	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR492/SRR492399/SRR492399.sra
7387	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR492/SRR492400/SRR492400.sra
7394	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR492/SRR492401/SRR492401.sra
7396	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945924
7404	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR492/SRR492402/SRR492402.sra
7411	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR492/SRR492406/SRR492406.sra
7415	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945928
7416	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR492/SRR492409/SRR492409.sra
7417	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945930
7419	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR492/SRR492241/SRR492241.sra
7424	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR492/SRR492289/SRR492289.sra
7427	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR492/SRR492302/SRR492302.sra
7430	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR492/SRR492328/SRR492328.sra
7460	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR492/SRR492240/SRR492240.sra
7471	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945938
7475	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945939
7477	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945940
7514	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR492/SRR492367/SRR492367.sra
7515	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR492/SRR492366/SRR492366.sra
7516	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR519/SRR519633/SRR519633.sra
7517	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR519/SRR519634/SRR519634.sra
7518	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR519/SRR519635/SRR519635.sra
7519	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR519/SRR519636/SRR519636.sra
7520	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR492/SRR492318/SRR492318.sra
7521	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR492/SRR492319/SRR492319.sra
7523	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945947
7525	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945948
7529	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945949
7530	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945950
7566	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945951
7568	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945952
7717	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945953
7757	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945954
7767	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945955
7917	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945956
7947	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945957
8037	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945958
8057	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945959
8077	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945960
8132	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945961

Supplementary Table 6: see page 134

Appendix

ID	URL
8171	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945962
8214	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR492/SRR492275/SRR492275.sra
8222	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR519/SRR519637/SRR519637.sra
8227	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR519/SRR519638/SRR519638.sra
8230	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR519/SRR519639/SRR519639.sra
8231	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR519/SRR519640/SRR519640.sra
8233	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR492/SRR492242/SRR492242.sra
8234	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR519/SRR519641/SRR519641.sra
8235	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR492/SRR492281/SRR492281.sra
8236	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR492/SRR492285/SRR492285.sra
8237	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR519/SRR519642/SRR519642.sra
8238	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR492/SRR492296/SRR492296.sra
8239	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945974
8240	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR519/SRR519643/SRR519643.sra
8241	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR519/SRR519644/SRR519644.sra
8242	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR519/SRR519645/SRR519645.sra
8243	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR492/SRR492342/SRR492342.sra
8244	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR492/SRR492343/SRR492343.sra
8246	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR492/SRR492329/SRR492329.sra
8247	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR519/SRR519646/SRR519646.sra
8249	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR519/SRR519647/SRR519647.sra
8256	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR519/SRR519648/SRR519648.sra
8258	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR519/SRR519649/SRR519649.sra
8259	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR519/SRR519650/SRR519650.sra
8264	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR492/SRR492219/SRR492219.sra
8283	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR519/SRR519651/SRR519651.sra
8284	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR492/SRR492248/SRR492248.sra
8285	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR492/SRR492249/SRR492249.sra
8290	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR492/SRR492256/SRR492256.sra
8297	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR492/SRR492267/SRR492267.sra
8306	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR519/SRR519652/SRR519652.sra
8307	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1945993
8311	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR492/SRR492286/SRR492286.sra
8312	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR492/SRR492287/SRR492287.sra
8326	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR519/SRR519653/SRR519653.sra
8334	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR519/SRR519654/SRR519654.sra
8335	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR519/SRR519655/SRR519655.sra
8337	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR492/SRR492322/SRR492322.sra
8343	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR492/SRR492327/SRR492327.sra
8351	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR519/SRR519656/SRR519656.sra
8354	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR492/SRR492341/SRR492341.sra
8357	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR492/SRR492345/SRR492345.sra
8365	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR492/SRR492355/SRR492355.sra
8366	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR492/SRR492356/SRR492356.sra
8369	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR519/SRR519657/SRR519657.sra
8376	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR519/SRR519658/SRR519658.sra
8386	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR519/SRR519659/SRR519659.sra
8387	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR519/SRR519660/SRR519660.sra
8419	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR492/SRR492405/SRR492405.sra
8420	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR492/SRR492295/SRR492295.sra
8422	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR519/SRR519661/SRR519661.sra
8424	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR492/SRR492293/SRR492293.sra
8426	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR519/SRR519662/SRR519662.sra
8427	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946015
8464	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946016
8483	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946017
8699	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946018
8723	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946019
9027	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946020
9057	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR519/SRR519663/SRR519663.sra
9058	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR519/SRR519664/SRR519664.sra
9067	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946023
9069	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946024
9070	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946025
9075	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946026

Supplementary Table 6: see page 134

ID	URL
9078	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946027
9079	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946028
9081	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946029
9084	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946030
9085	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946031
9089	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946032
9091	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946033
9095	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946034
9099	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946035
9100	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946036
9102	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946037
9103	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946038
9104	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946039
9105	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946040
9106	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946041
9111	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946042
9113	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946043
9114	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946044
9115	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946045
9121	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946046
9125	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946047
9128	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946048
9130	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946049
9131	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946050
9133	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946051
9134	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946052
9298	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946053
9314	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946055
9321	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR519/SRR519665/SRR519665.sra
9323	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR519/SRR519666/SRR519666.sra
9332	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR519/SRR519667/SRR519667.sra
9336	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR519/SRR519668/SRR519668.sra
9339	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR519/SRR519669/SRR519669.sra
9343	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946061
9352	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR519/SRR519670/SRR519670.sra
9353	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR519/SRR519671/SRR519671.sra
9363	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR519/SRR519672/SRR519672.sra
9369	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR519/SRR519673/SRR519673.sra
9370	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946066
9371	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR519/SRR519674/SRR519674.sra
9380	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR519/SRR519675/SRR519675.sra
9381	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR519/SRR519676/SRR519676.sra
9382	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR519/SRR519677/SRR519677.sra
9383	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR519/SRR519678/SRR519678.sra
9386	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR519/SRR519679/SRR519679.sra
9388	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR519/SRR519680/SRR519680.sra
9390	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR519/SRR519681/SRR519681.sra
9391	

Supplementary Table 6: see page 134

Appendix

ID	URL
9434	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR519/SRR519700/SRR519700.sra
9436	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR519/SRR519701/SRR519701.sra
9437	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR519/SRR519702/SRR519702.sra
9442	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR519/SRR519703/SRR519703.sra
9450	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR519/SRR519704/SRR519704.sra
9451	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR519/SRR519705/SRR519705.sra
9452	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR519/SRR519706/SRR519706.sra
9453	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR519/SRR519707/SRR519707.sra
9454	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR519/SRR519708/SRR519708.sra
9455	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR519/SRR519709/SRR519709.sra
9470	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR519/SRR519710/SRR519710.sra
9471	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946102
9471	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR519/SRR519711/SRR519711.sra
9476	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR519/SRR519712/SRR519712.sra
9481	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR519/SRR519713/SRR519713.sra
9503	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946105
9506	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946106
9507	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946107
9508	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946108
9509	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946109
9511	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946111
9512	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946112
9513	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946113
9514	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946114
9515	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946115
9517	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946116
9518	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946117
9519	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946118
9520	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946119
9521	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946120
9522	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946121
9523	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946122
9524	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946123
9525	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946124
9526	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946125
9527	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946126
9528	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946127
9529	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946128
9530	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946129
9531	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946130
9532	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946131
9533	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946132
9534	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946133
9535	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946134
9536	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946135
9537	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946136
9539	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946137
9540	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946138
9541	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946139
9542	https://sra-download.ncbi.nlm.nih.gov/srap

Supplementary Table 6: see page 134

[illegible]

Supplementary Table 6: see page 134

Appendix

[illegible]

Supplementary Table 6: see page 134

ID	URL
9709	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946287
9710	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946288
9711	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946289
9712	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946290
9713	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946291
9714	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946292
9716	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946293
9717	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946294
9718	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946295
9719	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946296
9720	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946297
9721	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946298
9722	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946299
9723	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946300
9725	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946301
9726	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946302
9727	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946303
9728	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946304
9729	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946305
9730	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946306
9731	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946307
9732	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946308
9733	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946309
9735	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946310
9736	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946311
9737	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946312
9738	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946313
9739	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946314
9741	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946315
9743	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946316
9744	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946317
9745	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946318
9747	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946319
9748	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946320
9749	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946321
9754	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946322
9755	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946323
9756	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946324
9757	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946325
9758	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR492/SRR492203/SRR492203.sra
9759	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR492/SRR492209/SRR492209.sra
9761	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR492/SRR492217/SRR492217.sra
9762	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR492/SRR492263/SRR492263.sra
9764	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR492/SRR492352/SRR492352.sra
9766	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR492/SRR492404/SRR492404.sra
9768	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946332
9769	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946333
9770	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946334
9771	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946335
9772	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946336
9774	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946337
9776	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946339
9777	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946340
9778	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946341
9780	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946343
9781	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946344
9782	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946345
9783	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946346
9784	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946347
9785	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946348
9786	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946349
9787	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946350
9788	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946351
9789	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946352
9790	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946353

Supplementary Table 6: see page 134

Appendix

[illegible]

Supplementary Table 6: see page 134

ID	URL
9859	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946419
9860	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946420
9861	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946421
9862	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946422
9864	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946423
9866	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946424
9867	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946425
9868	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946426
9869	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946427
9870	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946428
9871	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946429
9873	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946430
9874	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946431
9875	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946432
9876	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946433
9877	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946434
9878	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946435
9879	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946436
9880	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946437
9881	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946438
9882	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946439
9883	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946440
9885	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946441
9886	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946442
9887	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946443
9888	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946444
9890	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946445
9891	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946446
9892	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946447
9894	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946448
9895	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946449
9897	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946450
9898	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946451
9899	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946452
9900	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946453
9901	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946454
9902	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946455
9903	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946456
9904	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946457
9905	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946458
9906	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946459
9908	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946460
9910	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946462
9911	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946463
9912	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946464
9914	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946465
9915	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946466
9917	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946467
9918	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946468
9920	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946469
9921	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946470
9924	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946471
9925	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946472
9926	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946473
9927	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946474
9929	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946476
9930	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946477
9932	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946478
9933	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946479
9935	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946480
9937	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946481
9938	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946482
9939	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR095/SRR095761/SRR095761.sra
9939	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR095/SRR095762/SRR095762.sra
9941	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR095/SRR095695/SRR095695.sra

Supplementary Table 6: see page 134

[illegible]

132

[illegible]

Supplementary Table 6: see page 134

Appendix

ID	URL
10018	ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR095/SRR095760/SRR095760.sra
10020	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946551
10022	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946552
10023	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946553
10027	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946554
14312	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946555
14313	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946556
14314	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946557
14315	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946558
14318	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946559
14319	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946560
15560	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946561
15591	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946562
15592	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946563
15593	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946564
18694	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946565
18696	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946566
19949	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946567
19950	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946568
19951	https://sra-download.ncbi.nlm.nih.gov/srapub/SRR1946569

Supplementary Table 6: SRA files used in this study.

A total of 1055 sequences were analyzed. ID and URL correspond to the accession number and the link to download the SRA file, respectively.

Abbreviations

General

:	Fused to (in the context of gene fusion constructs)
::	Under the control of (in the context of promoter-gene constructs)
%	Percentage
♀	Female
♂	Male
®	Registered marked symbol
°C	Degree Celsius
3C	Chromatin conformation capture
3'	Three prime end of DNA fragment
5'	Five prime end of DNA fragment
μ	Micro
Agrobacterium	<i>Agrobacterium tumefaciens</i>
ANOVA	Analysis of variance
Arabidopsis	<i>Arabidopsis thaliana</i>
ASE	Allele-specific expression
BAM	Binary alignment map
BASTA	Glufosinate ammonium
bp	Base pair
bHLH	Basic helix-loop-helix
BT	Bolting time
C-	Carboxy terminal
cDNA	Complementary DNA
Chr	Chromosome
Col-0	Columbia
ChIP	Chromatin Immunoprecipitation
CRE	<i>Cis</i> -regulatory element
CORE	CONSTANS-responsive element
CRISPR-Cas	Clustered regularly interspaced short palindromic repeats/CRISPR-associated protein 9
DH	DNase I hypersensitivity

Abbreviations

DNA	Deoxyribonucleic acid
dNTP	Deoxyribonucleic triphosphate
Drosophila	<i>Drosophila melanogaster</i>
<i>E. coli</i>	<i>Escherichia coli</i>
EMSA	Electrophoresis mobility shift assay
enChIP	Engineered DNA-binding molecule ChIP
et al.	<i>et alii</i> / <i>et aliae</i> [Lat.] and others
F	Forward
F1, F2, F3	First, second, third... filial generation after a cross
FAST	Fluorescent-accumulating seed technology
g	Gram
GA	Gibberellic acid
GFP	Green fluorescent protein
GWAS	Genome-wide association study
h	Hour
HIF	Heterogeneous inbred family
H3K27me3	Tri-methylated lysine 27 at histone 3
InDel	Insertion-deletions
IL	Introgression line
IR	Inverted repeats
k	Kilo
kb	Kilobase pair
l	Liter
LD	Long-day
<i>Ler</i>	Landsberg <i>erecta</i>
lncRNA	Long noncoding RNA
LSF	Load Sharing Facility
M	Molar (mol/l)
m	Milli
MAF	Minor allele frequency
MAGIC	Multiparent advanced generation intercross
min	Minute
mol	Mol
mRNA	Messenger RNA

MULE	Mutator-like element
N-	Amino-terminal
n	Nano
NA	Not applicable
NIL	Near isogenic line
NCBI	National center for biotechnology information
NGS	Next generation sequencing
NOS	Nopaline synthase (<i>Agrobacterium tumefaciens</i>)
nsSNP	Nonsynonymous SNP
nt	Nucleotide
OD	Optical density
P	promoter
PcG	Polycomb group
PCR	Polymerase chain reaction
PEBP	Phosphatidyl ethanolamine binding domain protein
pH	Negative logarithm of proton concentration
PHD	Plant homeodomain
PI	Plastochron index
PRC	PcG repressive complex
Promoter	Sequence upstream of the TSS up to the upstream gene
qPCR	Quantitative PCR
QTL	Quantitative trait locus
R	Reverse
RIL	Recombinant inbred family
RNA	Ribonucleic acid
SA	Salicylic acid
SAM	Shoot apical meristem
SDs	Short day
SD	Standard deviation
SE	Standard error
SNP	Single nucleotide polymorphism
SRA	Short read archive
T1, T2, T3...	First, second, third... filial generation after a transformation
T-DNA	Transferred DNA

Abbreviations

TAIR	The Arabidopsis information resource
TE	Transposable element
TEP	Transposable element polymorphism
TFBS	Transcription factor binding site
TLN	Total leaf number
TSS	Translation start site
UTR	Untranslated region
VCF	Variant call format
Ws	Wassilewskija
x	Crossed to
ZT	Zeitgeber time

Genes and proteins

The nomenclature for plant genes follows the *Arabidopsis* standard. Genes are written in upper case italics. Mutants are indicated in lower case italics. Proteins appear in upper case regular letters.

35S	35S RNA promoter (<i>Cauliflower mosaic virus</i>)
AP (1-2)	APETALA (1-2)
CaMv	Cauliflower Mosaic Virus
CLF	CURLY LEAF
CO	CONSTANS
COL	CO-LIKE
COP1	CONSTITUTIVE PHOTOMORPHOGENESIS 1
E(Z)	ENHANCER OF ZESTE (<i>Drosophila melanogaster</i>)
FD	FLOWERING LOCUS D
FKF1	FLAVIN KELCH FACTOR 1
FLC	FLOWERING LOCUS
FLK	FLOWERING LOCUS KH DOMAIN
FLM	FLOWERING LOCUS M
FPA	FLOWERING LOCUS PA
FRI	FRIGIDA

FT	FLOWERING LOCUS T
FTIP1	FT-INTERACTING PROTEIN 1
FUL	FRUITFULL
FVE	FLOWERING LOCUS VE
FY	FLOWERING LOCUS Y
GI	GIGANTEA
GUS	β -glucoronidase (<i>Escherichia coli</i>)
LD	LUMINIDEPENDENS
LFY	LEAFY
LHP1	LIKE HETEROCHROMATIN PROTEIN1 (also TFL2)
Luc	Firefly luciferase (<i>Photinus pyralis</i>)
OLE1	OLEOSIN 1
MADS	MCM1 AGAMOUS DEFICIENS SRF
MEA	MEDEA
NF-Y (B-C)	NUCLEAR FACTOR Y (B-C)
miR156	microRNA156
miR172	microRNA172
NF-Y	NUCLEAR FACTOR-Y
PhyA	Phytochrome A
PhyB	Phytochrome B
PIF (4-5)	PHYTOCHROME-INTERACTING FACTOR (4-5)
PP2A	PROTEIN PHOSPHATASE 2
RKIP	RAF KINASE INHIBITOR PROTEIN
SMZ	SCHLAFMUTZE
SNZ	SCHNARCHZAPFEN
SOC1	SUPPRESSOR OF OVEREXPRESSION OF CONSTANS 1
SPA (1)	SUPPRESSOR OF PHYA (1)
SPL (1-15)	SQUAMOSA PROMOTER BINDING PROTEIN LIKE (1-15)
SU(Z)12	SUPPRESSOR OF ZESTE (<i>Drosophila melanogaster</i>)
SVP	SHORT VEGETATIVE PHASE
SWN	SWINGER
TEM (1-2)	TEMPRANILLO (1-2)
TFL2	TERMINAL FLOWER 2 (also LHP1)
TOE (1-3)	TARGET OF EAT (1-3)

Abbreviations

TSF	TWIN SISTER OF FT
VIN3	VERNALIZATION INSENSITIVE 3
VRN5	VERNALIZATION 5

Amino acids

Alanine	A
Cysteine	C
Aspartic acid	D
Glutamic acid	E
Phenylalanine	F
Glycine	G
Histidine	H
Isoleucine	I
Lysine	K
Leucine	L
Methionine	M
Asparagine	N
Proline	P
Glutamine	Q
Arginine	R
Serine	S
Threonine	T
Valine	V
Tryptophane	W

Acknowledgements

I would like to thank...

Dr. Franziska Turck for her supervision, her support and patience even when I had the weirdest theories. These four years were just great. Thank you so much.

Prof. Dr. George Coupland for being my first examiner and for his support.

Prof. Dr. Ute Höcker for being my second examiner.

Prof. Dr. Maria Albani for being the chair of my examination committee.

My Ph.D. supervisors: **Dr. José Jimenez-Gomez**, **Dr. Fabian Bratzel** and **Dr. Angela Hancock**.

All the people who also gave me advice, supports and inspirations during this Ph.D.: **Dr. Guillaume Née**, **Dr. Amaury de Montaigne**, **Dr. Korbinian Schneeberger**, **Dr. Ding Jia**, **Dr. Mathieu Piednoel**, **Dr. Artem Pankin**, **Dr. Fernando Andrés**, **Dr. Stefan Wötzel** and **Dr. Jörg Wunder**.

The former and actual Turkies, especially **Petra Tänzler** for her great support, **Dr. Yue Zhou** for being my supervisor when I arrived here, **Dr. Johan Zicola** for the scientific discussions and **Alexander Förderer** for being a great friend.

All the people I had the chance to meet here: **Dr. Theodoros Zografou**, **Dr. Samson Simon**, **Dr. Maida Romerat-Branchat** and **Dr. Matthias Berens**.

The people who convinced me to do this Ph.D.: **Dr. Cristel Carles** especially, **Dr. Robert Blanvillain**, **Dr. Christine Ondzighi-Assoume** and **Dr. Gabriela Bucini**.

The **gardeners** for taking care of my plants, being so friendly and helping me to practice my German.

The **IT service**, for their help, even when I was close to crash the server.

My **friends** Polochon, Cake, Niaki, Benji-san, Carotte, Crevette, Tunisiano, Bouba, Abdel, Laurane, Minmin, Greg, Hans, Glinglin, Elodie, Belette, Aiekick, Julie, Berthe, Clementin, Bastal, Colombix, Golly... I am sorry we became more distant recently, but to be honest, I would not have been able to do this Ph.D. with you around me... ☺ I love you!

Merci à tout' mon côté rényoné, lo **pèr**, mon deux **matantes**, mon ban couzinsn couzines, bana lé super! Mi aim zot toute!

Merci a ma **famille**, mes **grand parents**, **Payou**.

Maman, un grand Merci pour tout ce que tu fais pour moi. Je t'aime!

And last but not least, meine Liebe **Karinnen**, for being my muse and strongest support every day.

Erklärung

Ich versichere, dass ich die von mir vorgelegte Dissertation selbständig angefertigt, die benutzen Quellen und Hilfsmittel vollständig angegeben und die Stellen der Arbeit einschließlich Tabellen, Karten und Abbildungen, die anderen Werken im Wortlaut oder dem Sinn nach entnommen sind, in jedem Einzelfall als Entlehnung kenntlich gemacht habe; dass diese Dissertation noch keiner anderen Fakultät oder Universität zur Prüfung vorgelegen hat; dass sie - abgesehen von unten angegebenen Teilpublikationen – noch nicht veröffentlicht worden ist sowie, dass ich eine solche Veröffentlichung vor Abschluss des Promotionsverfahrens nicht vornehmen werde. Die Bestimmungen dieser Promotionsordnung sind mir bekannt. Die von mir vorgelegte Dissertation ist von Prof. Dr. George Coupland betreut worden.

Köln, July 2018 _____
Emmanuel Tergemina

Teilpublikationen

Es liegen keine Teilpublikationen vor.

Lebenslauf

Persönliche Angaben

Name: Emmanuel Tergemina
Geburtsdatum: 30 Mai 1986
Geburtsort: Saint Martin d'Hères
Nationalität: Französisch
Familienstand: Ledig

Ausbildung

Seit 2013

Promotions-Studium an der Universität zu Köln, Köln, Deutschland

Max-Planck-Institut für Züchtungsforschung

Stipendium der International Max-Planck-Research School

2010-2013

International Master in Molekular und Zellular Biologie an der Universität Grenoble

Alpes, Grenoble, Frankreich

2006-2010

Bachelor in Biologie an der Universität Grenoble Alpes, Grenoble, Frankreich

2004-2006

Preparatory year of Medicin an der Universität Claude Bernard, Lyon, Frankreich

2004

Baccalaureat, Lycee Pierre du Terrail, Pontcharra, Frankreich

Köln, July 2018